Data depth

Pavlo Mozharovskyi*

*LTCI, Telecom Paris, Institut Polytechnique de Paris

Tail events analysis:
Robustness, outliers and models for extreme values

Palaiseau, February 18, 2020

Outline of the course

Format: 6 $\times$ 3.5 hours + exam

- Class 1: Introduction to robust statistics
- Class 2: Lab session I
- Class 3: Data depth
- Weeks 4: Extreme value statistics
- Week 5 : Multi-dimensional setting
- Week 6: Lab session II

Programming language: `R`

Grading: Exam

Today

The concept of data depth

Some examples
- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth
- Properties
- Trimmed regions

Classes of depth functions
- Depths satisfying the projection property

A concept of depth for the functional framework
- Functional outlier detection

References

Contents

The concept of data depth

Some examples
- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth
- Properties
- Trimmed regions

Classes of depth functions
- Depths satisfying the projection property

A concept of depth for the functional framework
- Functional outlier detection

References

Statistical data depth (Zuo & Serfling, 2000)

A data depth measures how "close" a given point is located to the "center" of a distribution. For $\boldsymbol{x} \in \mathbb{R}^d$ and a d-variate random vector X distributed as $P \in \mathcal{P}$, a data depth is a function

$$D : \mathbb{R}^d \times \mathcal{P} \to \mathbb{R}^1, (\boldsymbol{x}, P) \mapsto D(\boldsymbol{x}|P)$$

that is:

(i) affine invariant: $D(A\boldsymbol{x} + b|AX + b) = D(\boldsymbol{x}|X)$ for any $d \times d$ non-singular matrix A and any $b \in \mathbb{R}^d$;

(ii) maximal at symmetry center: $D(\boldsymbol{x}^*|X) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}|X)$ holds for any $P \in \mathcal{P}$ having center $\boldsymbol{x}^*$;

(iii) monotone on rays: for any $\boldsymbol{x}^*$ with $D(\boldsymbol{x}^*|X) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}|X)$, any $\boldsymbol{x} \in \mathbb{R}^d$, and any $0 \leq \alpha \leq 1$ it holds: $D(\boldsymbol{x}|X) \leq D(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)|X)$;

(iv) vanishing at infinity: $\lim_{||\boldsymbol{x}|| \to \infty} D(\boldsymbol{x}|X) = 0$.

A distribution is called centrally symmetric about $\boldsymbol{x}^*$ if distributions of $X - \boldsymbol{x}^*$ and $\boldsymbol{x}^* - X$ coincide.

Statistical data depth (Mosler, 2013)

A data depth measures how "close" a given point is located to the "center" of a distribution. For $\boldsymbol{x} \in \mathbb{R}^d$ and a d-variate random vector X distributed as $P \in \mathcal{P}$, a data depth is a function

$$D : \mathbb{R}^d \times \mathcal{P} \to [0, 1], (\boldsymbol{x}, P) \mapsto D(\boldsymbol{x}|P)$$

that is:

- D1 translation invariant: $D(\boldsymbol{x} + b|X + b) = D(\boldsymbol{x}|X)$ for any $b \in \mathbb{R}^d$;
- D2 linear invariant: $D(A\boldsymbol{x}|AX) = D(\boldsymbol{x}|X)$ for any $d \times d$ non-singular matrix A;
- D3 vanishing at infinity: $\lim_{||\boldsymbol{x}|| \to \infty} D(\boldsymbol{x}|X) = 0$;
- D4 monotone on rays: for any $\boldsymbol{x}^* \in \arg\max_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}|X)$, any $\boldsymbol{x} \in \mathbb{R}^d$, and any $0 \leq \alpha \leq 1$ it holds: $D(\boldsymbol{x}|X) \leq D(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)|X)$;
- D5 upper semicontinuous in $\boldsymbol{x}$: the upper-level sets $D_\alpha(X) = \{\boldsymbol{x} \in \mathbb{R}^d : D(\boldsymbol{x}|X) \geq \alpha\}$ are closed for all α.

Statistical data depth (Mosler, 2013)

Some remarks:

- D1 – D5 implies (ii) due to Dyckerhoff (2004):
  If X is centrally symmetric about $\boldsymbol{x}^*$, then any depth function satisfying D1 – D5 $D(\cdot|X)$ is maximal at $\boldsymbol{x}^*$.

- D4 implies star-shaped upper-level sets of D.

  One can strengthen to:

  - D4con: $D(\cdot|X)$ is a quasiconcave function, *i.e.* the upper-level sets $D_\alpha(X)$ are convex for all α.

- D1 and D2 define affine invariante depth.

  One can also weaken to:

  - D2iso: $D(A\boldsymbol{x}|AX) = D(\boldsymbol{x}|X)$ for every isometric linear A to define orthogonal invariant depth;

  - D2sca: $D(\lambda\boldsymbol{x}|\lambda X) = D(\boldsymbol{x}|X)$ for any $\lambda > 0$ to define scale invariant depth.

Central regions

- For given P and $\alpha \in [0, 1]$, the level sets $D_\alpha(P)$ form a family of depth-trimmed of central regions.
- The innermost region arises at some depth $\alpha_{\max} \leq 1$, which depends on the depth notion D and distribution P. Then $D_\alpha(X)$ is the set of deepest points.
- Central regions describe distribution w.r.t. location, dispersion, and shape.
- Properties of central regions, for any α:
  - Due to D1 and D2 $D_\alpha(X)$ is affine equivariant: $D_\alpha(AX + b) = AD_\alpha(X) + b$ for any $d \times d$ non-singular matrix A and any $b \in \mathbb{R}^d$;
  - Due to D3 $D_\alpha(X)$ is bounded;
  - Due to D4 $D_\alpha(X)$-s are nested: if $\alpha \geq \beta$, then $D_\alpha(X) \subseteq D_\beta(X)$, and star-shaped; due to D4con $D_\alpha(X)$ is in addition convex;
  - Due to D5 $D_\alpha(X)$ is closed.

Depth lifts and orderings

- Assume $\alpha_{\max} = 1$ and add a real dimension to D_α, $\alpha \in [0,1]$, then construct the depth lift:

$$\hat{D}(P) = \left\{(\alpha, y) \in [0,1] \times \mathbb{R}^d \,:\, y = \alpha x,\, x \in D_\alpha(P),\, \alpha \in [0,1]\right\}.$$

- This gives rise to an ordering of probability distributions in $\mathcal{P}$:

$$P \prec_D Q \quad \text{if} \quad \hat{D}(P) \subset \hat{D}(Q)\,.$$

  If (and only if) the family of central regions induced by D completely characterizes the underlying distribution, then $\prec_D$ is and order, otherwise it is a preorder.

- D introduces a probability semi-metric on $\mathcal{P}$:

$$\delta_D(P, Q) = \delta_H\big(\hat{D}(P), \hat{D}(Q)\big)\,,$$

  where $\delta_H(X, Y)$ is the Hausdorff distance between two non-empty sets in $\mathbb{R}^d$ with metric δ:

$$\delta_H(X, Y) = \max\Big\{\sup_{x \in X} \inf_{y \in Y} \delta(x, y),\, \sup_{y \in Y} \inf_{x \in X} \delta(x, y)\Big\}\,.$$

Applications

- Multivariate data analysis (Liu, Parelius, Singh '99);
- Statistical quality control (Liu, Singh '93);
- Cluster analysis and classification (Mosler, Hoberg '06; Li, Cuesta-Albertos, Liu '12; Lange, Mosler, Mozharovskyi '14);
- Tests for multivariate location, scale, symmetry (Liu '92; Dyckerhoff '02; Dyckerhoff, Ley, Paindaveine '15);
- Outlier detection (Hubert, Rousseeuw, Segaert '15);
- Multivariate risk measurement (Cascos, Mochalov '07);
- Robust linear programming (Bazovkin, Mosler '15);
- Missing data imputation (Mozharovskyi, Josse, Husson '17);
- etc.

R-package `ddalpha` (Pokotylo, Mozharovskyi, Dyckerhoff, Nagy): calculates a number of depths; performs depth-based classification of multivariate and functional data; contains 50 multivariate and 5 functional data sets.

Contents

The concept of data depth

Some examples
- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth
- Properties
- Trimmed regions

Classes of depth functions
- Depths satisfying the projection property

A concept of depth for the functional framework
- Functional outlier detection

References

Contents

The concept of data depth

Some examples

- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth

- Properties
- Trimmed regions

Classes of depth functions

- Depths satisfying the projection property

A concept of depth for the functional framework

- Functional outlier detection

References

Mahalanobis depth (Mahalanobis, 1936)

- Mahalanobis depth is defined as:

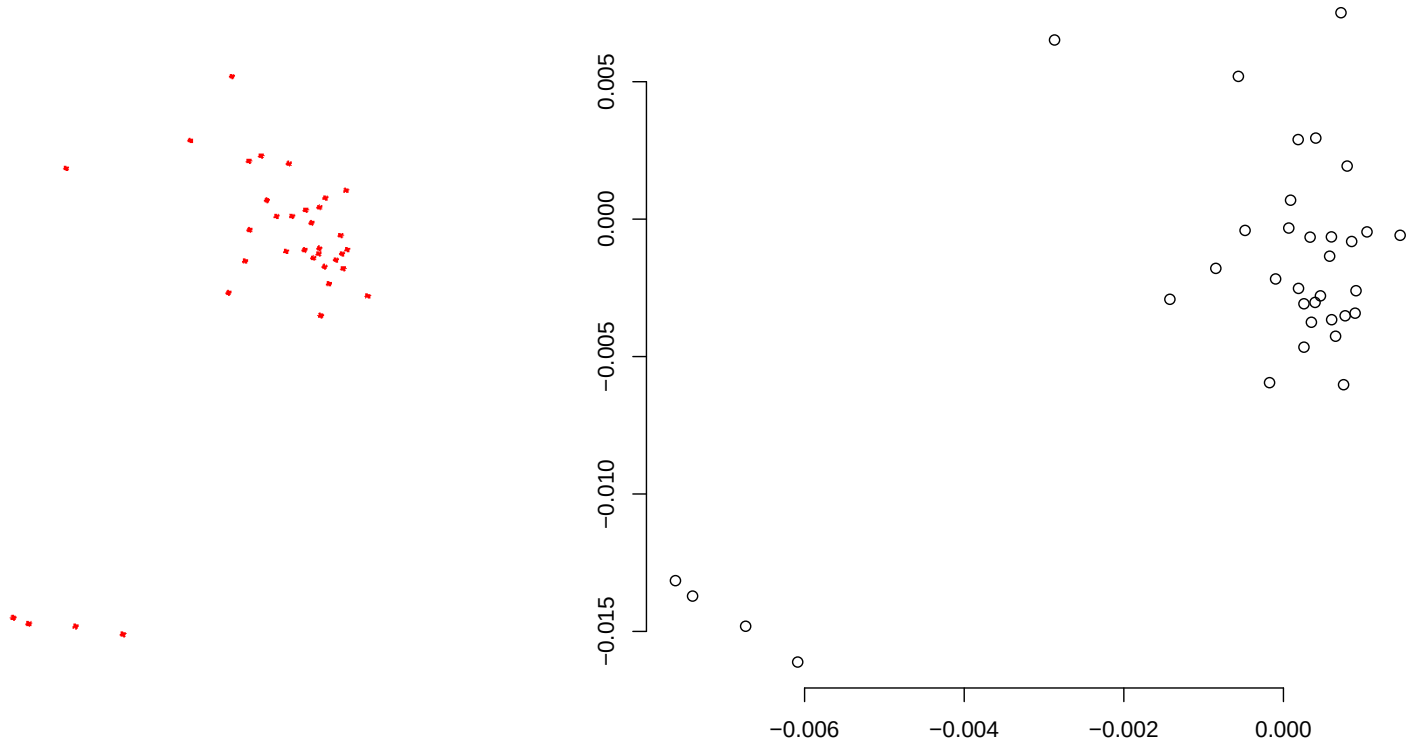
$$D^{Mah}(\boldsymbol{x}|X) = \frac{1}{1 + (d^{Mah})^2(\boldsymbol{x}|X)},$$

based on Mahalanobis distance:

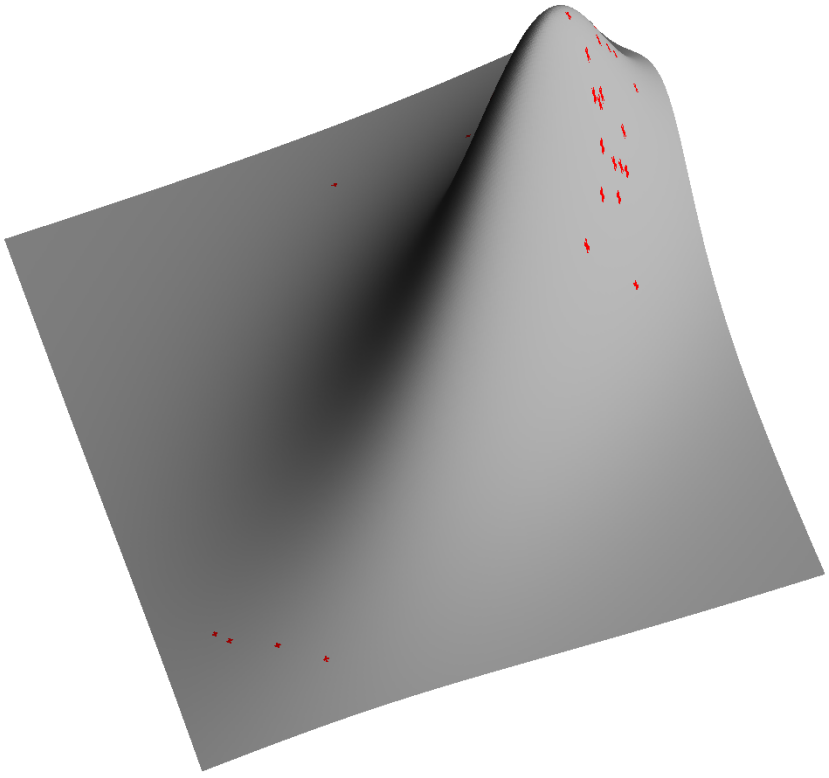
$$(d^{Mah})^2(\boldsymbol{x}|X) = (\boldsymbol{x} - \boldsymbol{\mu}_X)^T \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_X).$$

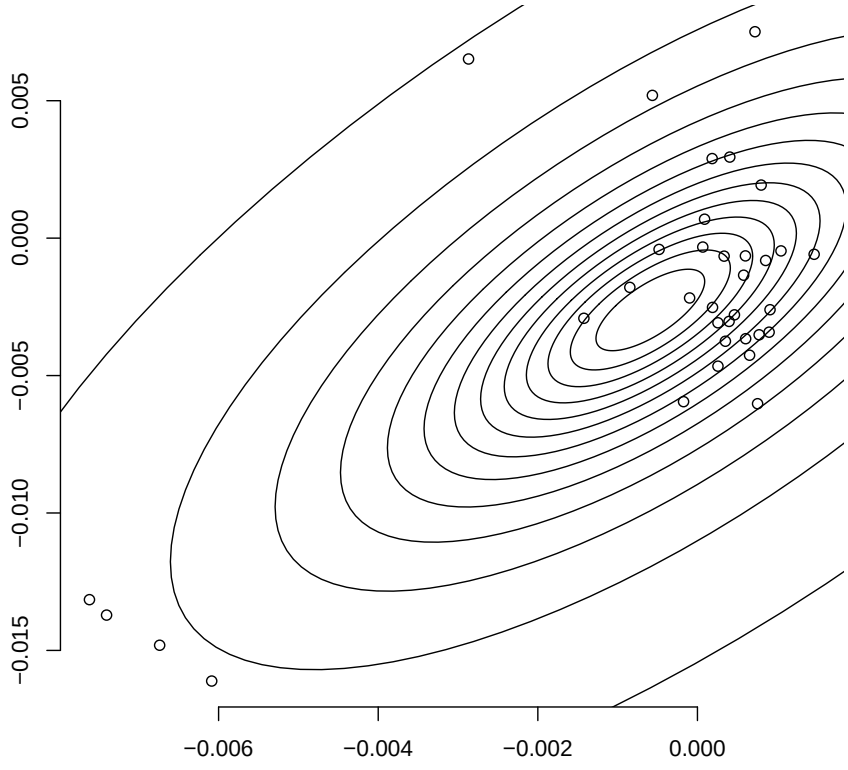
- In the empirical version, $\boldsymbol{\mu}_X$ and $\boldsymbol{\Sigma}_X$ are substituted by suitable estimates:
  - moment estimates;
  - robust estimates such as minimum volume ellipsoid or minimum covariance determinant (MCD).
- Properties:
  - satisfies D1 – D5 and D4con, is continuous;
  - being defined by $d(d+1)$ parameters, can be seen as a parametric depth;
  - by a single elliptical contour characterizes a multivariate normal distribution or one within an affine family of non-degenerate elliptical distributions.

Mahalanobis depth (Mahalanobis, 1936)



Mahalanobis depth (Mahalanobis, 1936)





Contents

The concept of data depth

Some examples

- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth

- Properties
- Trimmed regions

Classes of depth functions

- Depths satisfying the projection property

A concept of depth for the functional framework

- Functional outlier detection

References

Spatial depth (Vardi & Zhang, 2000; Serfling 2002)

Exploiting the idea of spatial quantiles of Chaudhuri (1996) and Koltchinskii (1997), Vardi & Zhang (2000) and Serflig (2002) formulate the spatial depth (also L_1-depth) as:

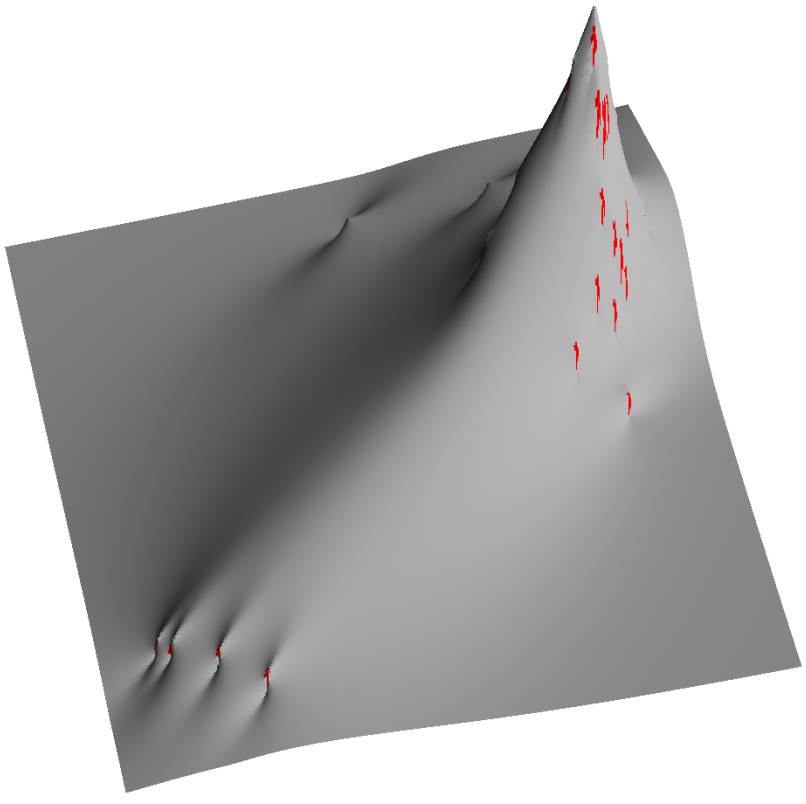
$$D^{spt}(\boldsymbol{x}|X) = 1 - \left\| \mathbb{E}\left[v\left(\boldsymbol{\Sigma}^{-\frac{1}{2}} (\boldsymbol{x} - X) \right) \right] \right\|, \quad v(\boldsymbol{y}) = \begin{cases} \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} & \text{if } \boldsymbol{y} \neq \boldsymbol{0}, \\ \boldsymbol{0} & \text{if } \boldsymbol{y} = \boldsymbol{0}. \end{cases}$$

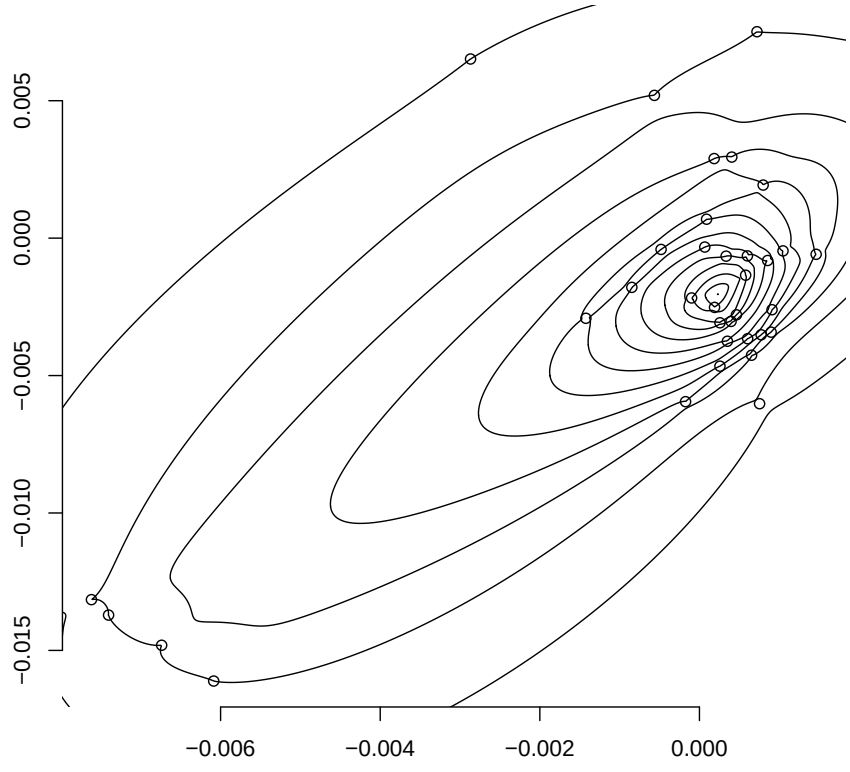
Properties:

- satisfies D1 – D5, but not D4con, is continuous;
- if $\boldsymbol{\Sigma}$ is orthogonal, satisfies D2iso only;
- with D2iso its maximum (say $\boldsymbol{x}^*$) is referred to as spatial median, a multivariate location estimator having asymptotic breakdown point of 0.5;
- spatial sign covariance matrix can be defined as:

$$S(X) = \mathbb{E}\left[v(X - \boldsymbol{x}^*) v(X - \boldsymbol{x}^*)^T \right].$$

Spatial depth (Vardi & Zhang, 2000; Serfling 2002)





Contents

The concept of data depth

Some examples

- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth

- Properties
- Trimmed regions

Classes of depth functions

- Depths satisfying the projection property

A concept of depth for the functional framework

- Functional outlier detection

References

Projection depth (Zuo & Serfling, 2000)

According to Zuo & Serfling (2000), projection depth is defined as:

$$D^{prj}(\boldsymbol{x}|X) = \frac{1}{1 + O^{prj}(\boldsymbol{x}|X)},$$

where

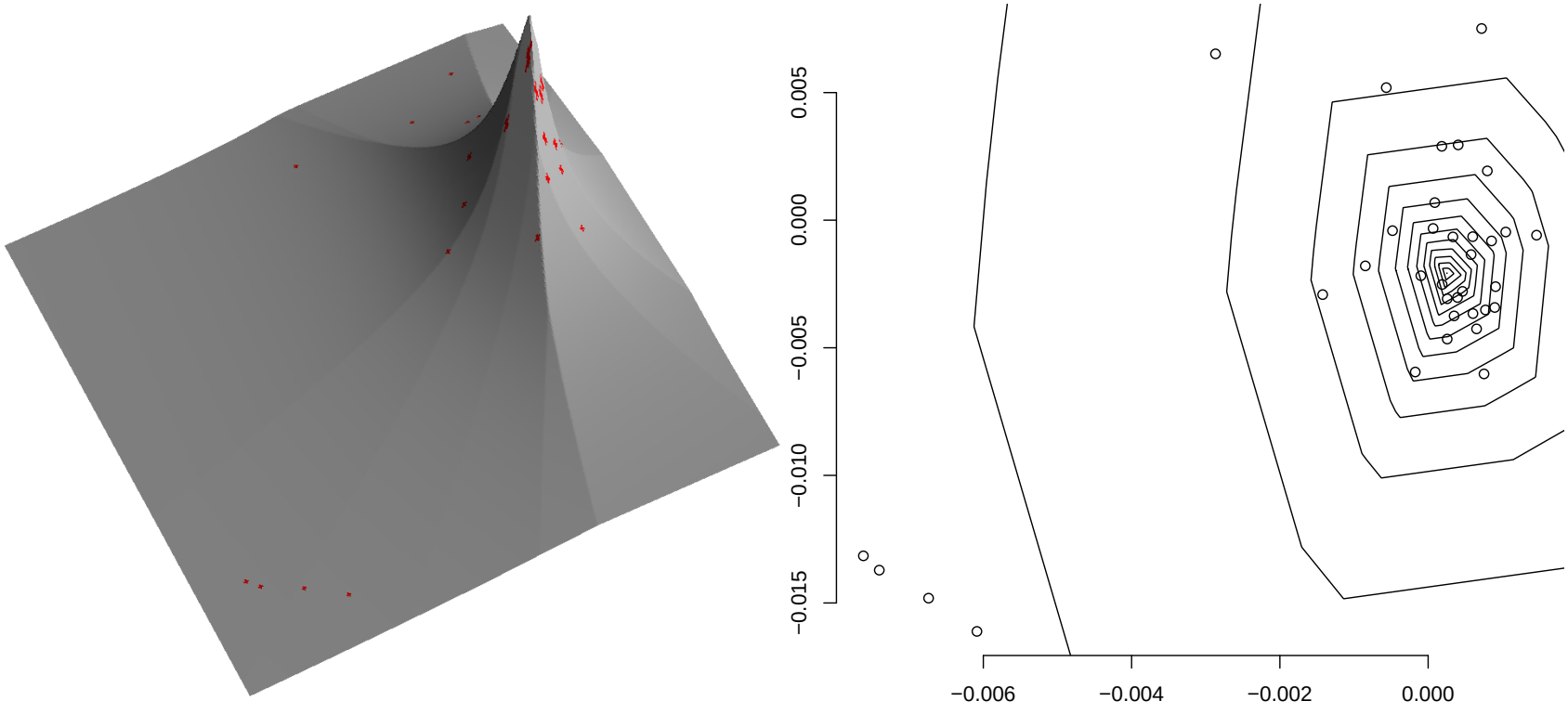
$$O^{prj}(\boldsymbol{x}|X) = \sup_{\boldsymbol{r} \in S^{d-1}} \frac{|\boldsymbol{x}^T \boldsymbol{r} - \text{med}(X^T \boldsymbol{r})|}{\text{MAD}(X^T \boldsymbol{r})}$$

is the projected outlyingness (Stahel, 1981; Donoho, 1982), $\text{med}(Y)$ and $\text{MAD}(Y) = \text{med}(|Y - \text{med}(Y)|)$ are the univariate median and median absolute deviation from the median, respectively.

Properties:

- Satisfies D1 – D5 and D4con, is continuous;
- its median has asymptotic breakdown point of 0.5.

Projection depth (Zuo & Serfling, 2000)



Contents

- The concept of data depth
- Some examples
  - Mahalanobis depth
  - Spatial depth
  - Projection depth
  - Zonoid depth
- Depth-based outlier detection
- Halfspace depth
  - Properties
  - Trimmed regions
- Classes of depth functions
  - Depths satisfying the projection property
- A concept of depth for the functional framework
  - Functional outlier detection
- References

Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

Koshevoy & Mosler (1997) and Mosler (2002) define the zonoid trimmed region as:

$$D_\alpha^{zon}(P) = \Big\{ \int_{\mathbb{R}^d} \mathbf{x} g(\mathbf{x}) dP(\mathbf{x}) : g : \mathbb{R}^d \mapsto \left[0, \frac{1}{\alpha}\right] \text{ measurable and } \int_{\mathbb{R}^d} g(\mathbf{x}) dP(\mathbf{x}) = 1 \Big\}$$

for $\alpha \in (0, 1]$, and

$$D_0^{zon}(P) = \mathrm{cl}\Big(\bigcup_{\alpha \in (0,1]} D_\alpha(P) \Big)$$

for $\alpha = 0$, where "cl" denotes closure.

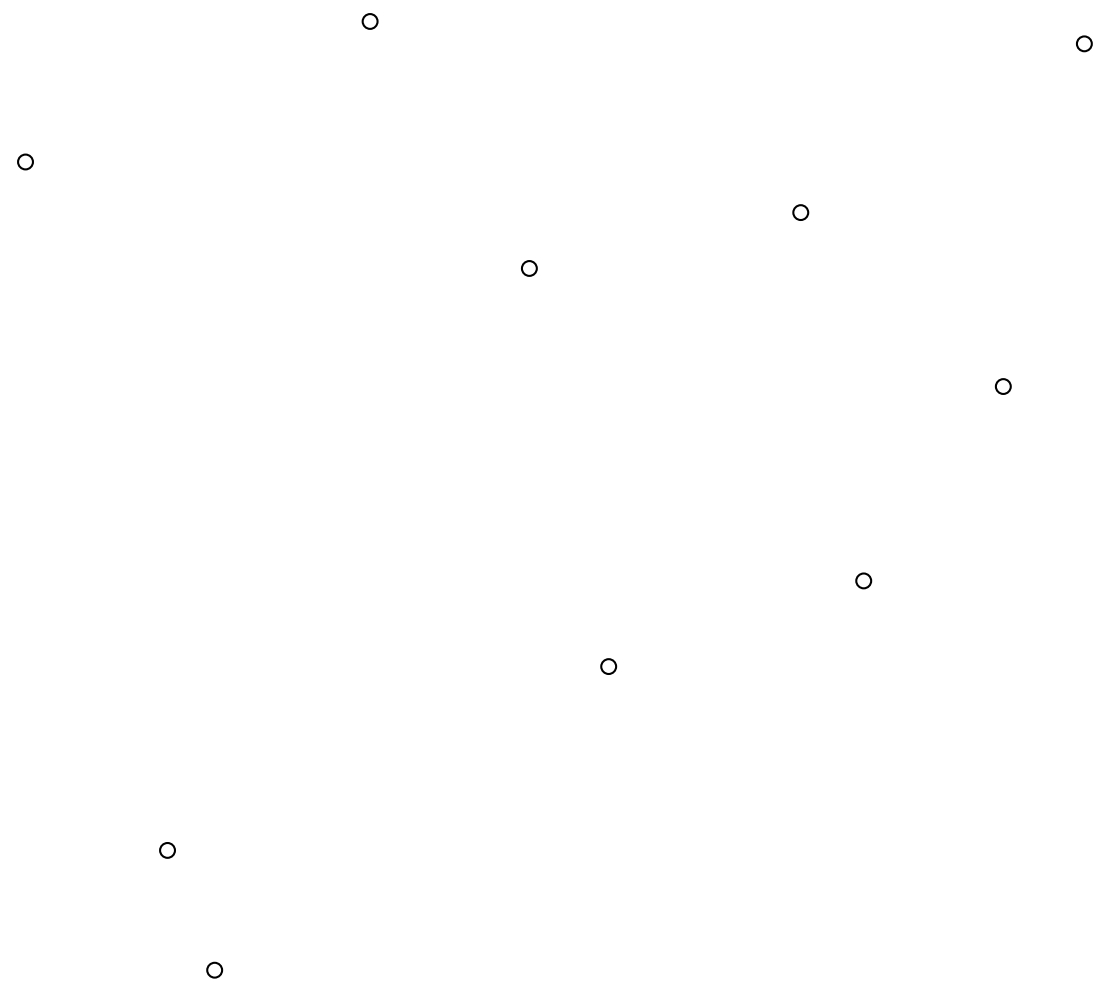
Zonoid depth is then defined as:

$$D^{zon}(\boldsymbol{x}|X) = \begin{cases} \sup\{\alpha : \boldsymbol{x} \in D_\alpha^{zon}(X)\} & \text{if } \boldsymbol{x} \in \mathrm{conv}\big(\mathrm{supp}(X)\big), \\ 0 & \text{otherwise}. \end{cases}$$

Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

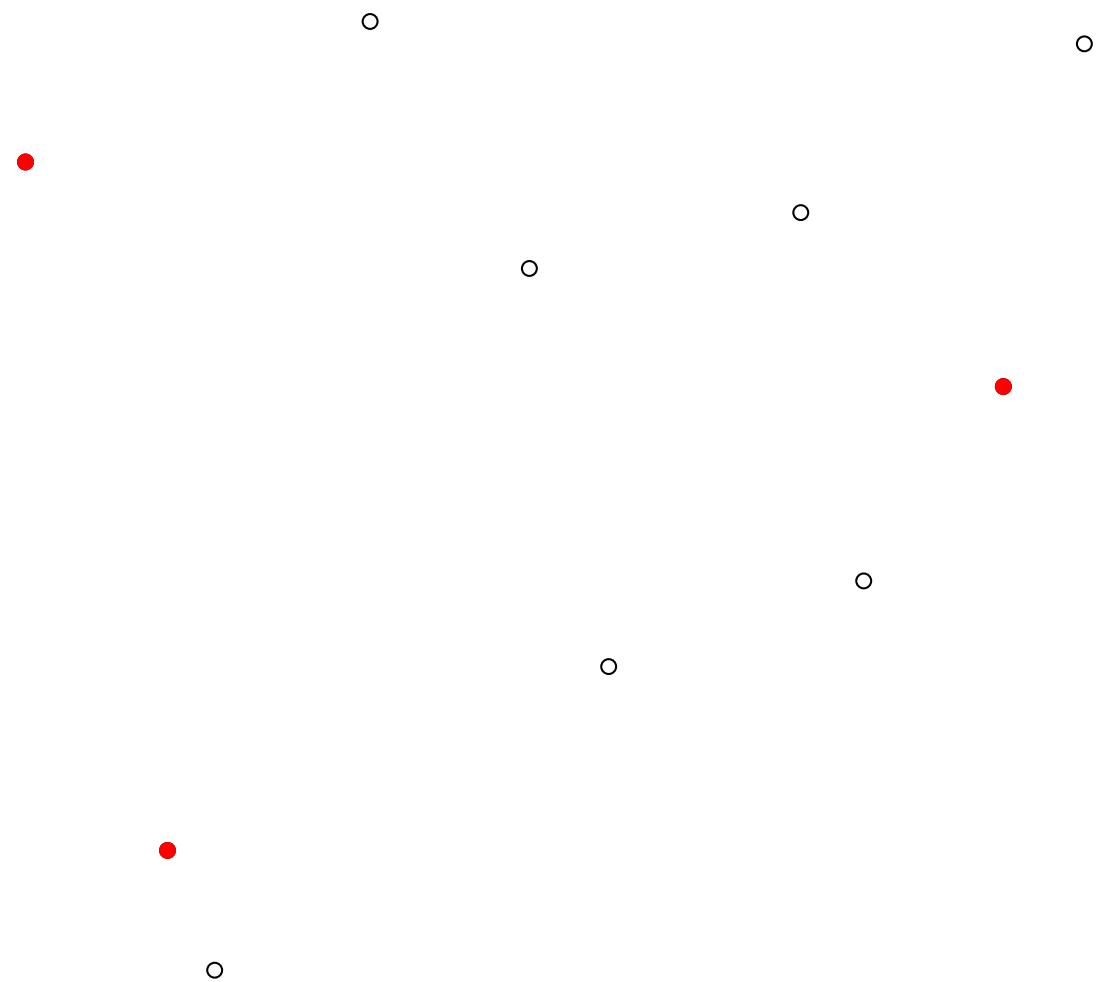
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

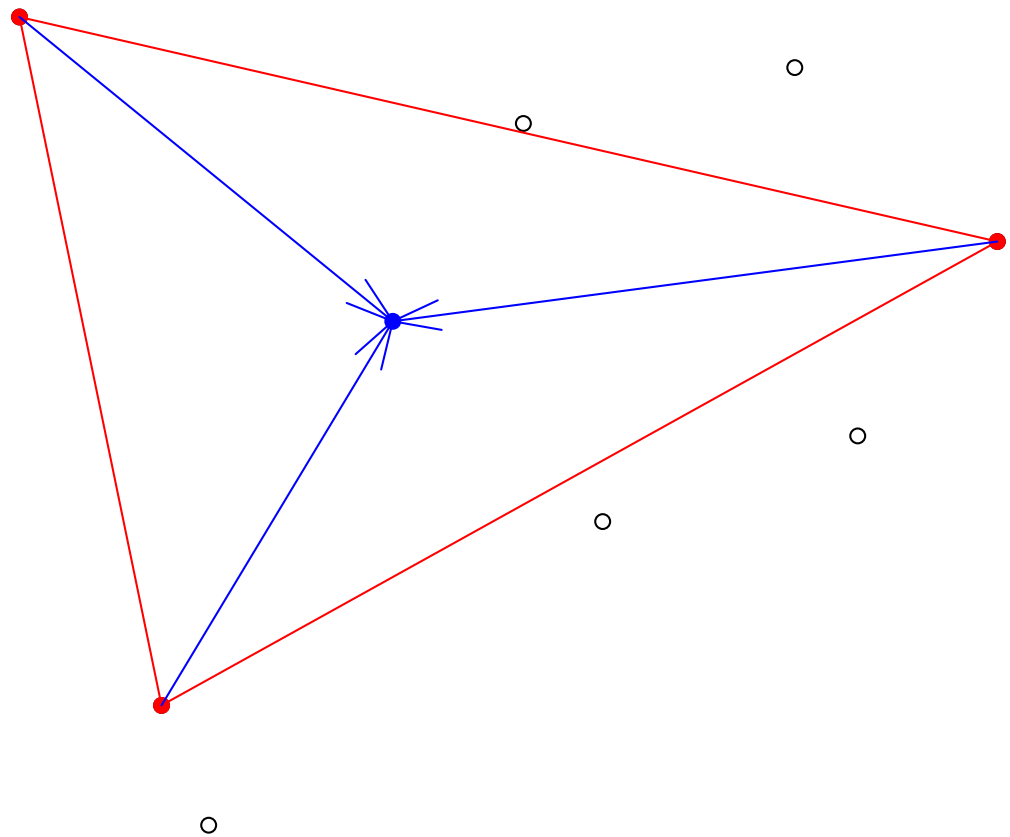
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

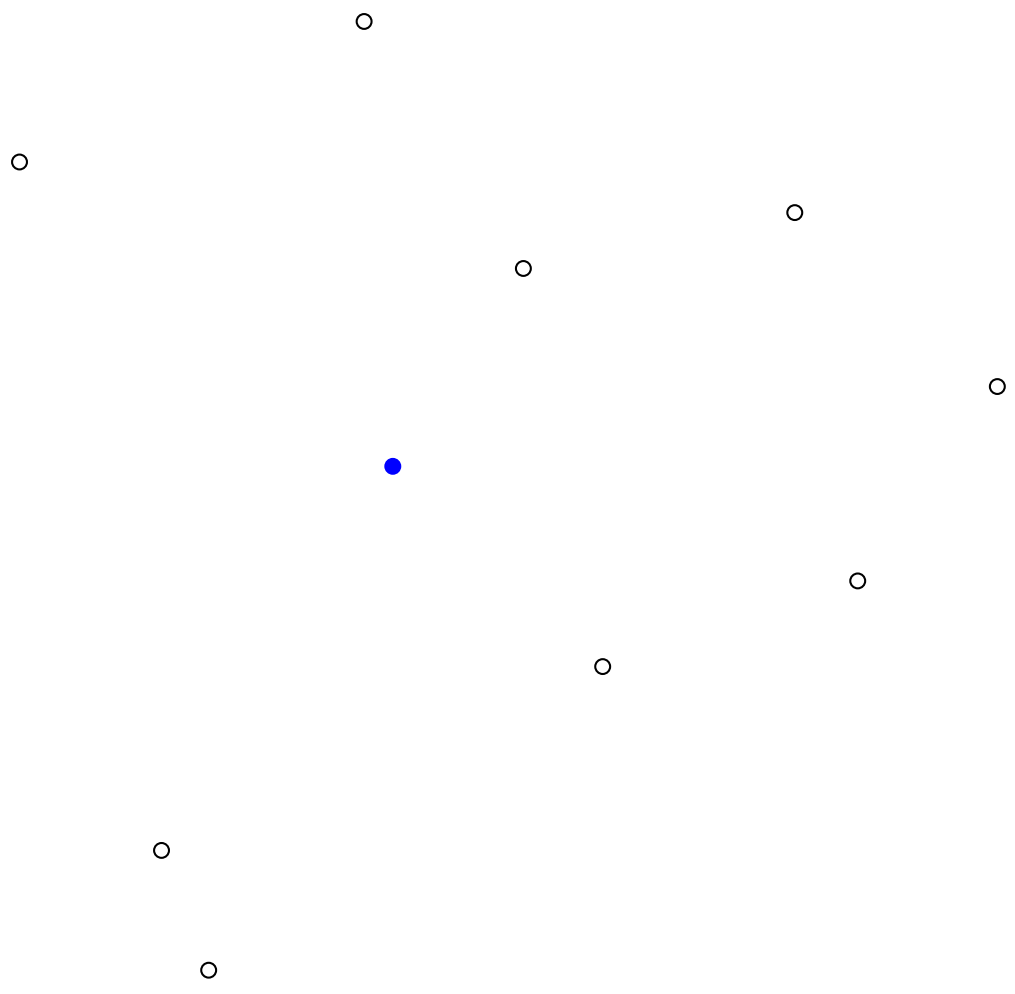
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

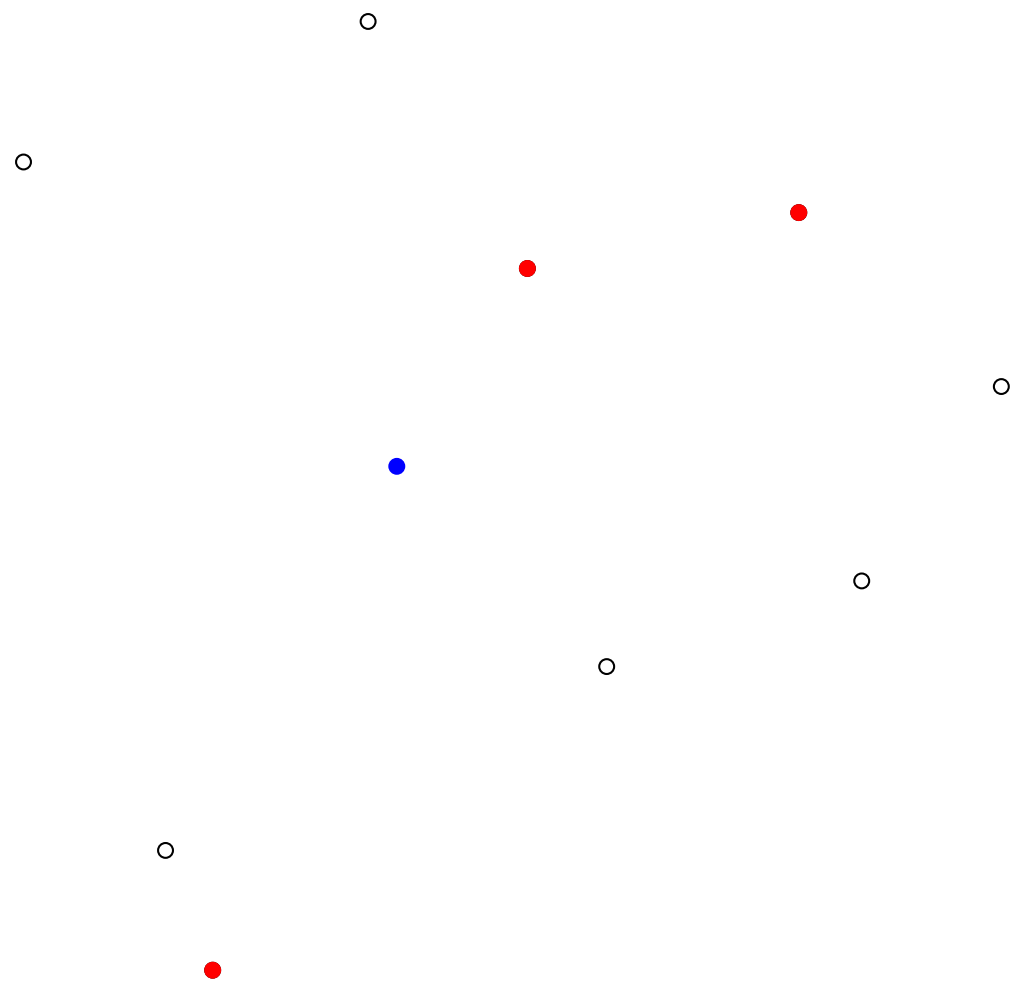
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

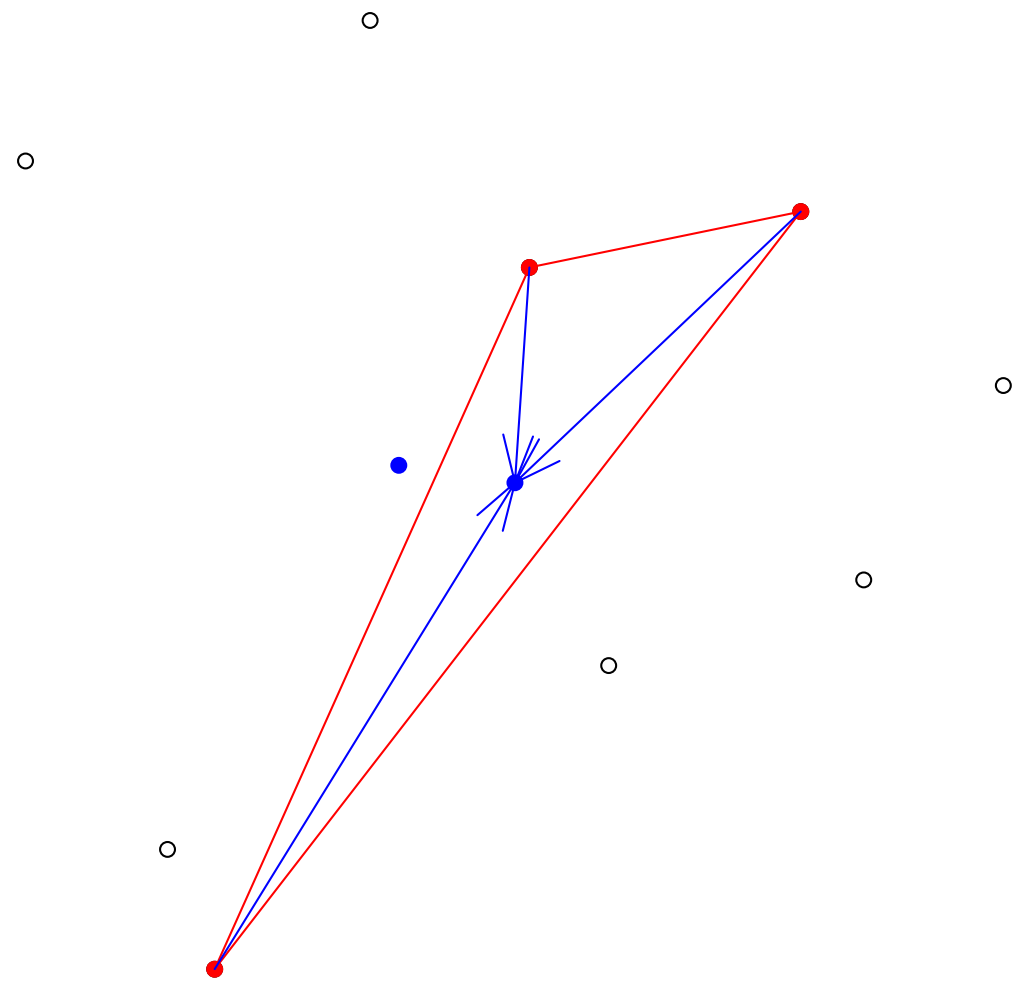
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

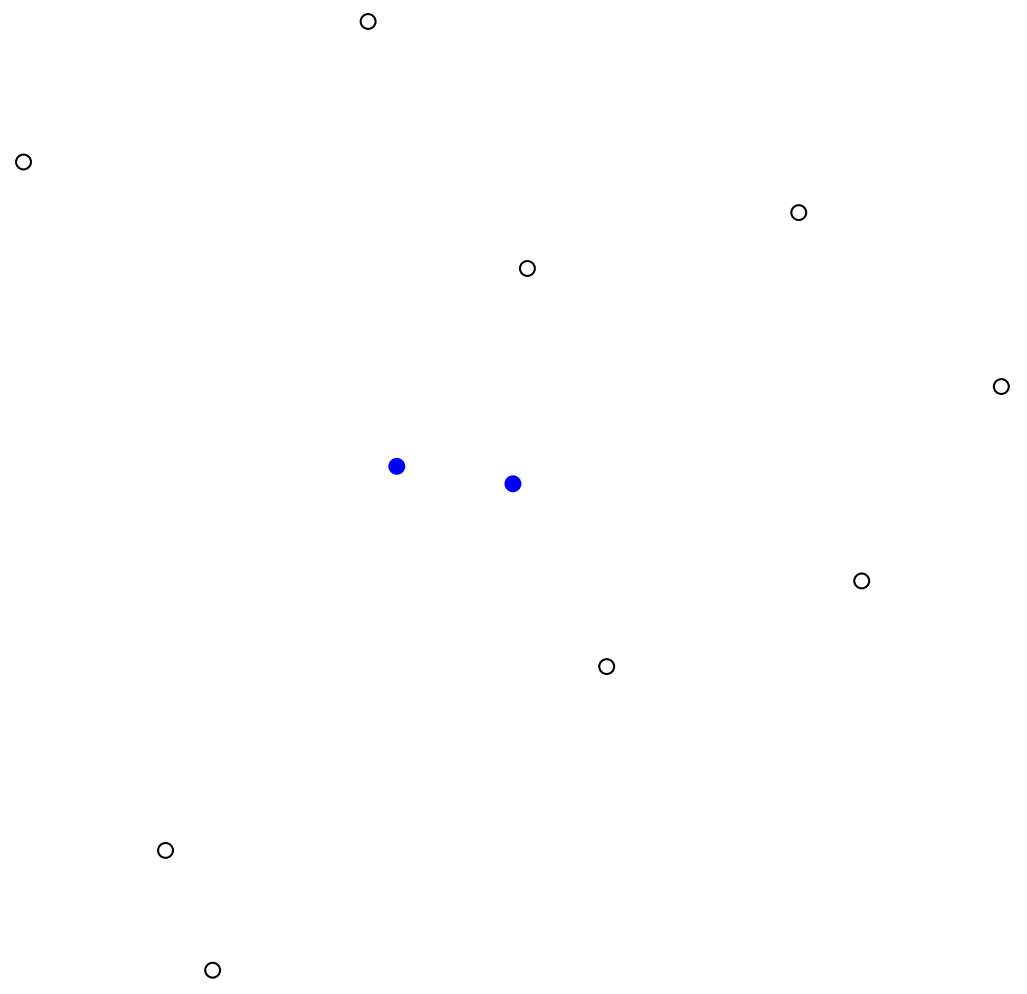
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

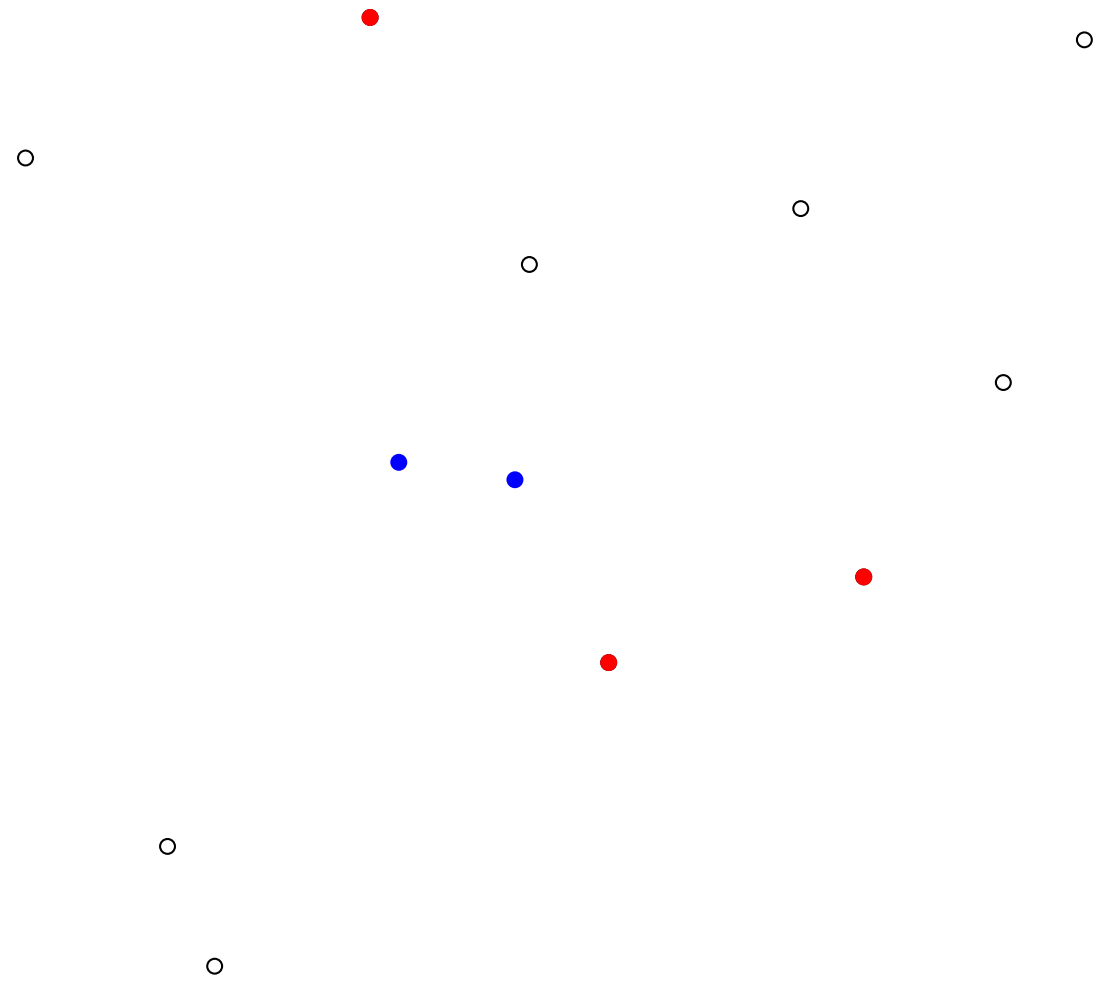
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

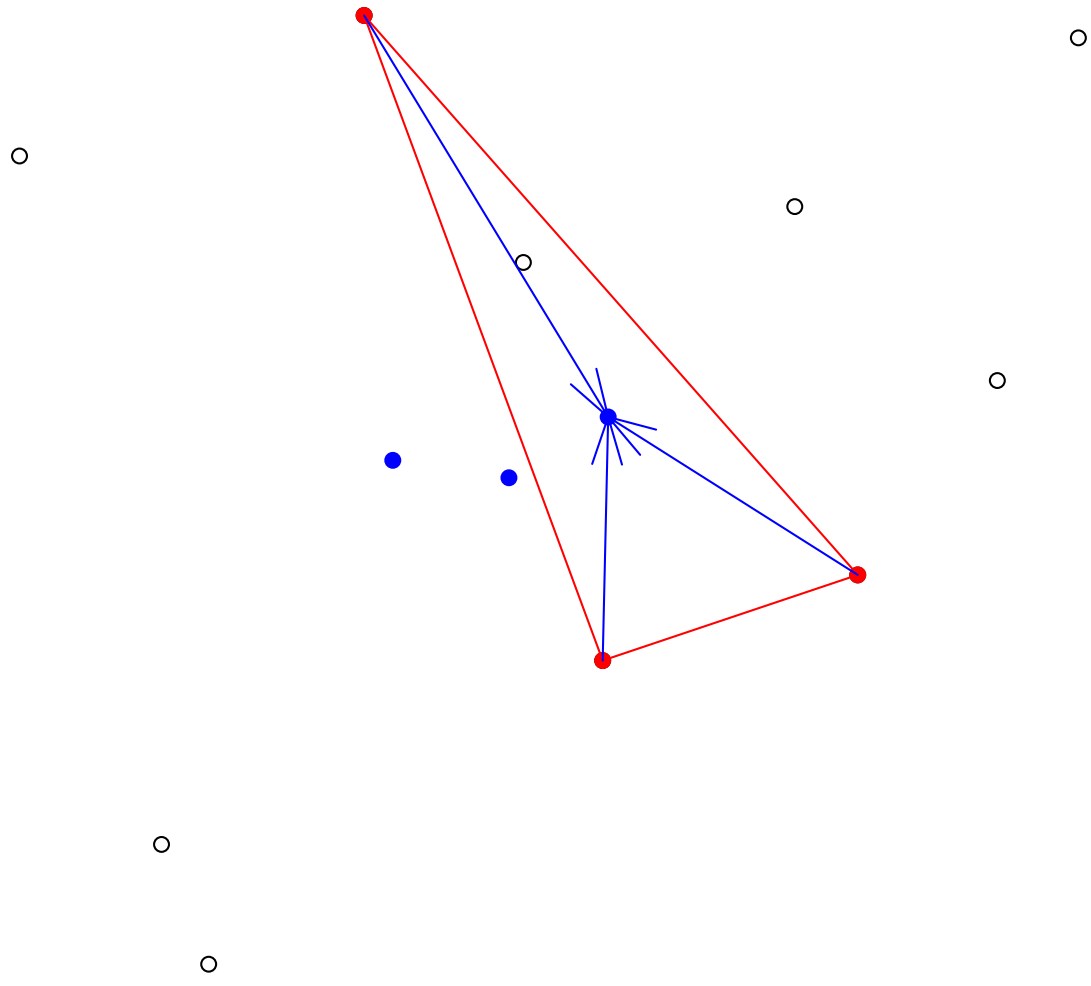
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

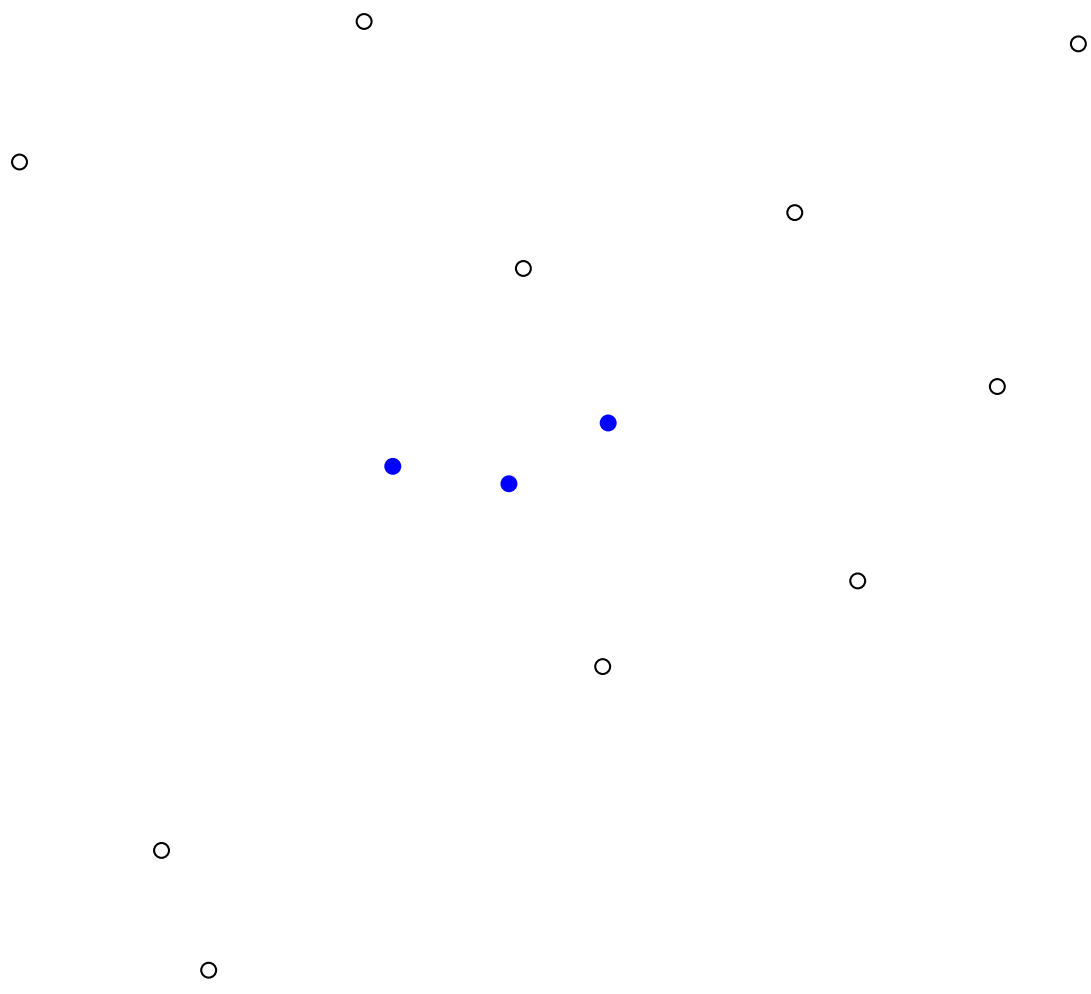
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

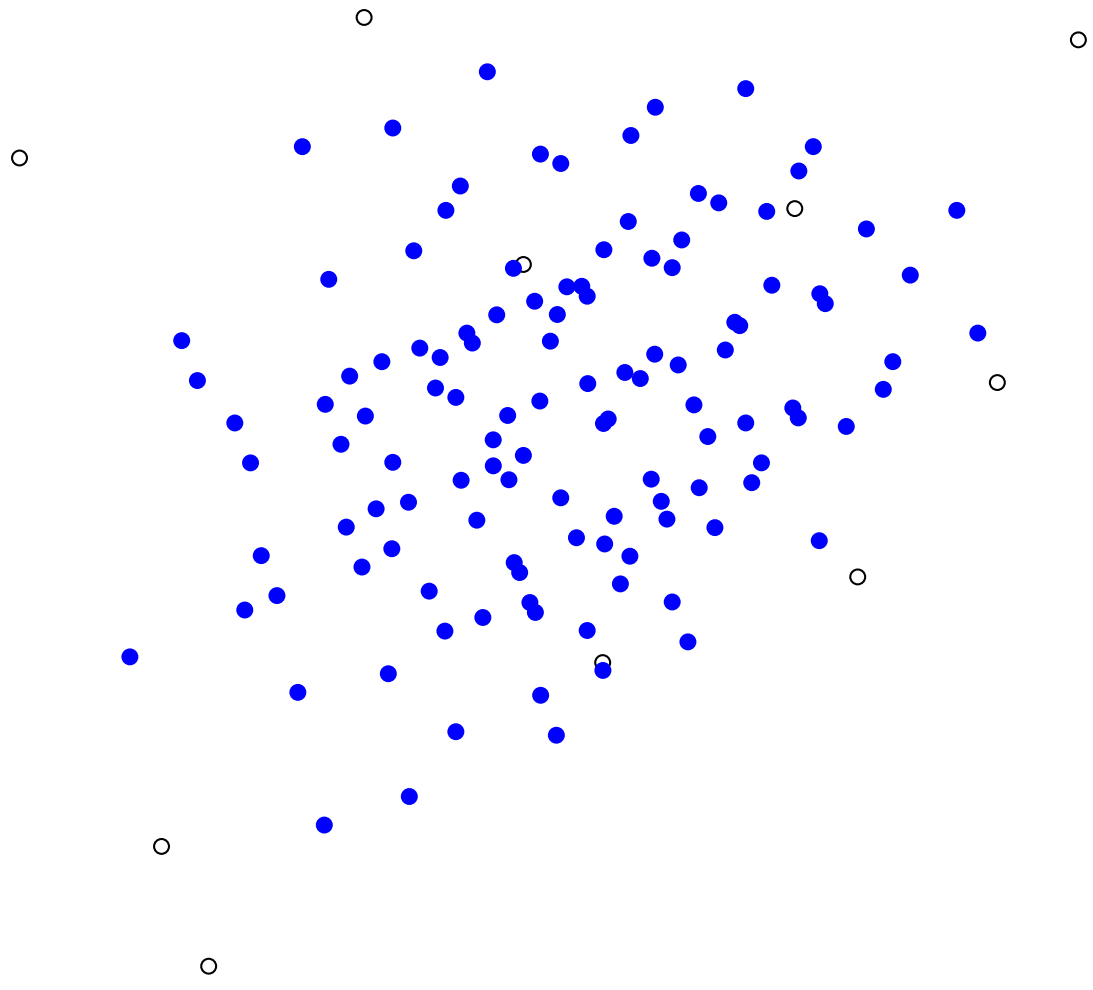
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

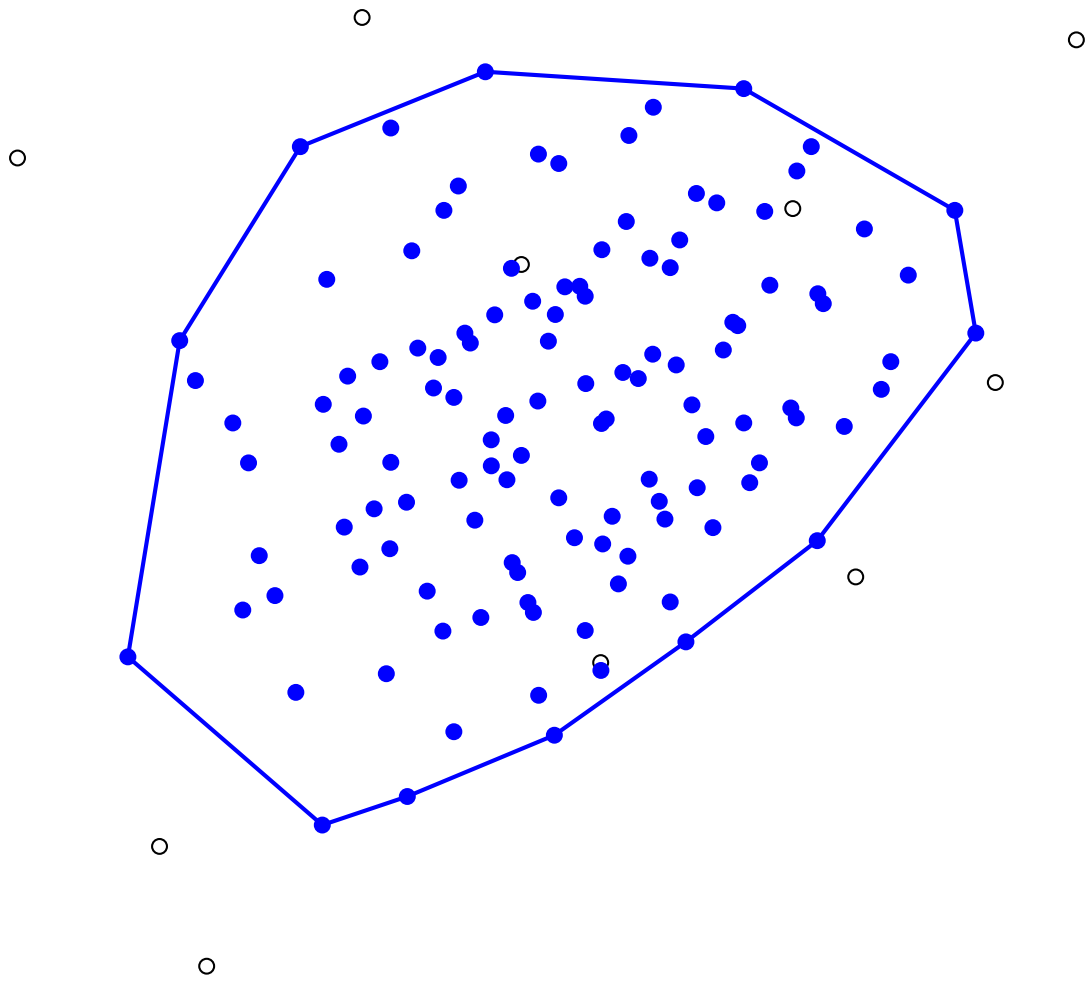
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

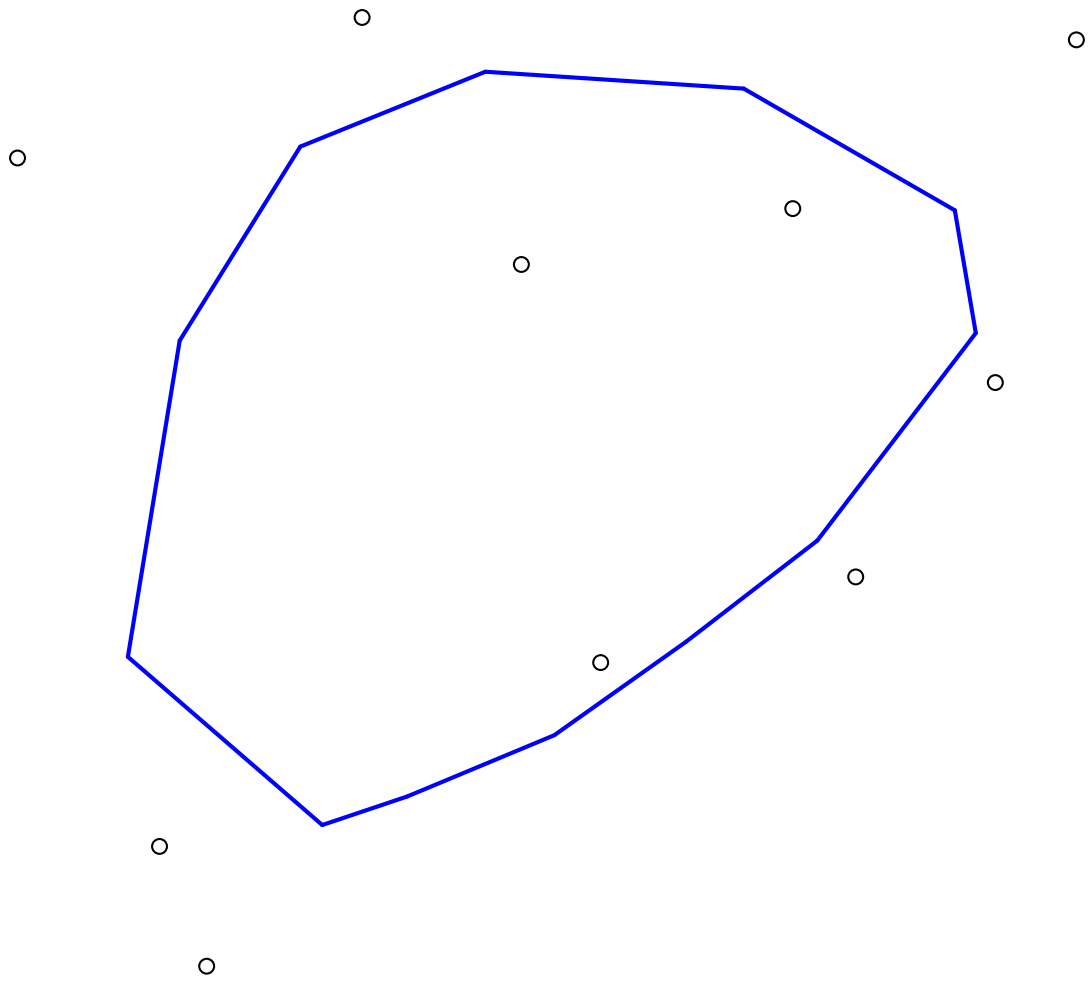
$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



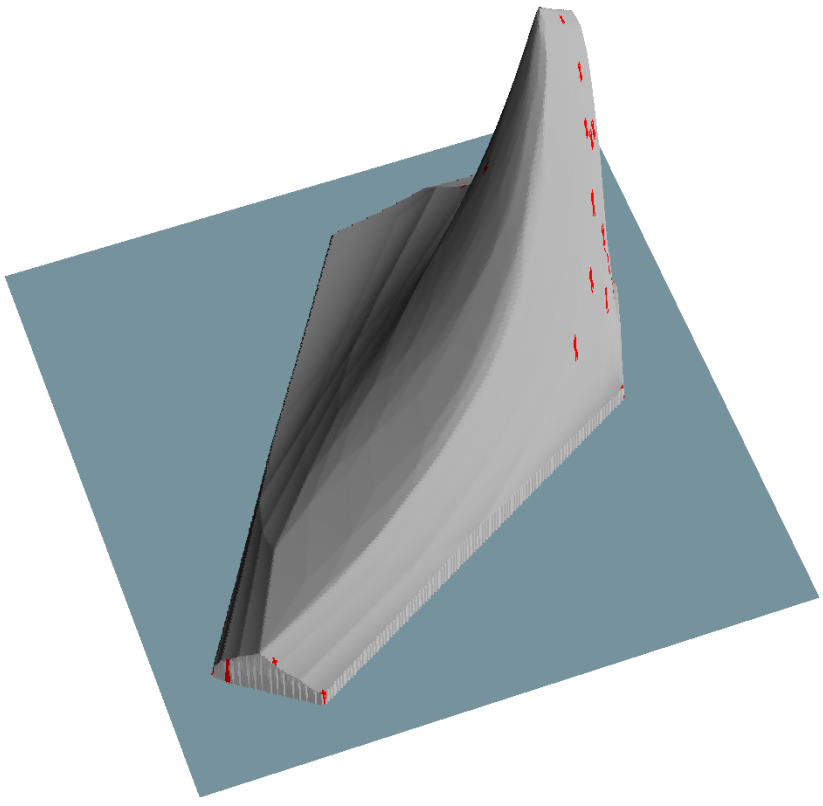
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

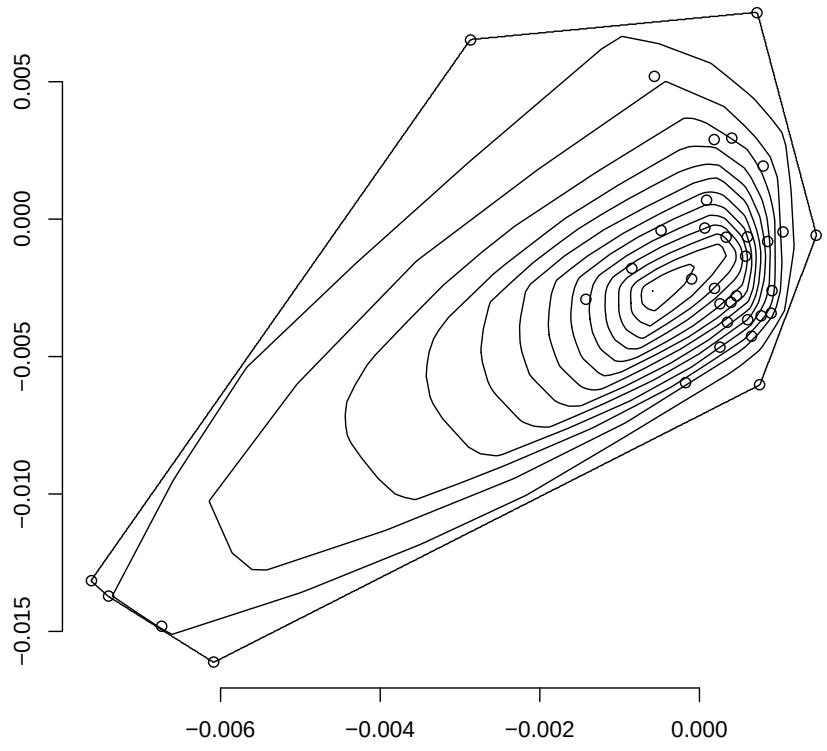
For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

$$D_\alpha^{zon}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)





Zonoid depth (Mosler, 2002), properties:

- satisfies D1 – D5 and D4con;
- $D^{zon}_{\alpha}(X)$ is equivariant to any linear transformation, *i.a.* any marginal projection of D^{zon}_{α} is D^{zon}_{α} of the marginal;
- $D^{zon}(\boldsymbol{x}|X)$ is continuous on $\mathrm{conv}\big(\mathrm{supp}(X)\big)$ in $\boldsymbol{x}$ and in X;
- $D^{zon}(\boldsymbol{x}|X)$ is monotone on dilation, *i.e.* for two random vectors X and Y in $\mathbb{R}^d$:

$$D^{zon}(\boldsymbol{x}|X) \leq D^{zon}(\boldsymbol{x}|Y) \quad \text{if } Y \stackrel{d}{=} \lambda X \text{ with } \lambda > 1\,;$$

- $D^{zon}(\boldsymbol{x}|X) = 0$ for all $\boldsymbol{x} \notin \mathrm{conv}\big(\mathrm{supp}(X)\big)$;
- $D^{zon}(\cdot|X)$ characterizes the distribution;
- $D^{zon}_{\alpha}(X)$-s are subadditive, *i.e.* for X and Y in $\mathbb{R}^d$:

$$D^{zon}_{\alpha}(X+Y) \subset D^{zon}_{\alpha}(X) \oplus D^{zon}_{\alpha}(Y)$$

  where $\oplus$ signifies the Minkowski sum of sets;
- If P has finite first moment and $P(\partial H) = 0$ for every halfspace H with $P(H) = 1$, then (Cascos, Lopez-Diaz, 2016):

$$\sup_{\boldsymbol{x}\in\mathbb{R}^d} \left| D^{zon}(\boldsymbol{x}|P) - D^{zon}(\boldsymbol{x}|P_n) \right| \xrightarrow{a.s.} 0\,.$$

Zonoid depth (Dyckerhoff, Koshevoy, Mosler, 1996):

- Zonoid depth can be defined as:

$$D^{zon}(\boldsymbol{x}|\boldsymbol{X}) = \sup\{\alpha \in [0,1] : \boldsymbol{x} \in D_\alpha^{zon}(\boldsymbol{X})\},$$

where sup of $\emptyset$ is defined to be 0, with:

$$D_\alpha^{zon}(\boldsymbol{X}) = \left\{\sum_{i=1}^n \lambda_i \boldsymbol{x}_i \ : \ \sum_{i=1}^n \lambda_i = 1, 0 \le \lambda_i, \alpha\lambda_i \le \frac{1}{n} \text{ for all } i\right\}.$$

- $D^{zon}(\boldsymbol{x}|\boldsymbol{X}) = \frac{1}{n\gamma^*}$ with γ^* being an optimizer of:

$$\begin{aligned} \min \gamma \quad \text{s. t.} \quad \boldsymbol{X}^T\boldsymbol{\lambda} &= \boldsymbol{x}, \\ \boldsymbol{\lambda}^T\mathbf{1}_n &= 1, \\ \gamma\mathbf{1}_n - \boldsymbol{\lambda} &\ge \mathbf{0}_n, \\ \boldsymbol{\lambda} &\ge \mathbf{0}_n. \end{aligned}$$

(Notation: $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^T$, $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)^T$.)

Contents

The concept of data depth

Some examples
- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth
- Properties
- Trimmed regions

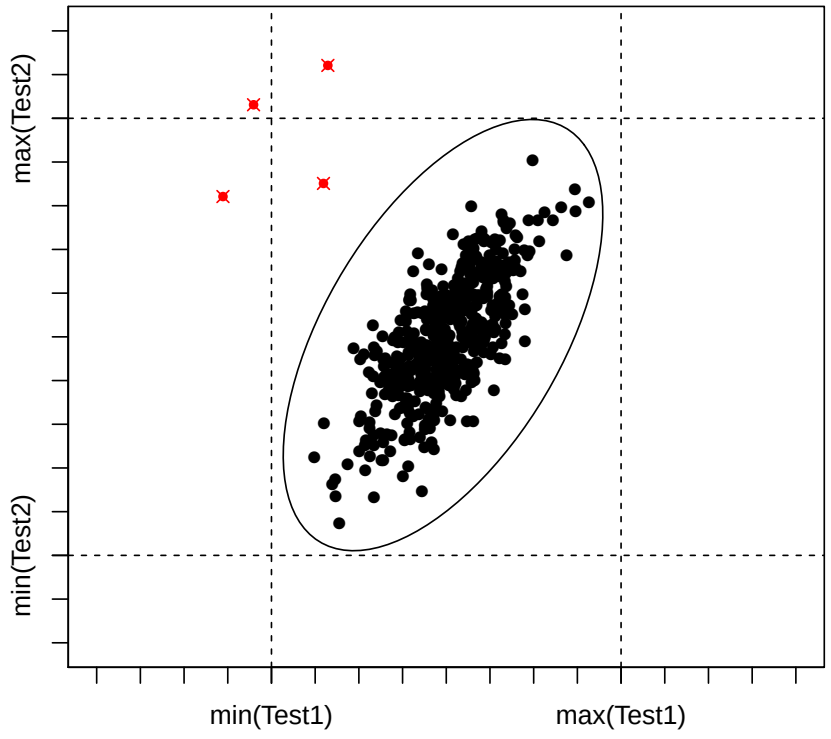
Classes of depth functions
- Depths satisfying the projection property

A concept of depth for the functional framework
- Functional outlier detection

References

Detection of multivariate outliers: Mahalanobis depth

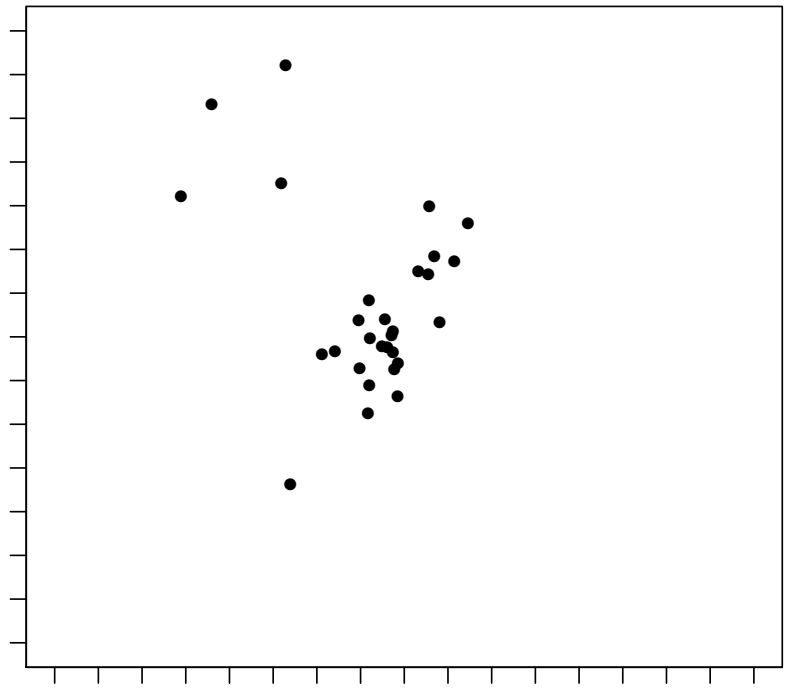
Regard again the two measurements during a test:



- Label $\boldsymbol{x}$ as outlier $D^{Mah}(\boldsymbol{x}|\boldsymbol{X}) < \min(D^{Mah})$.

Detection of multivariate outliers: Mahalanobis depth

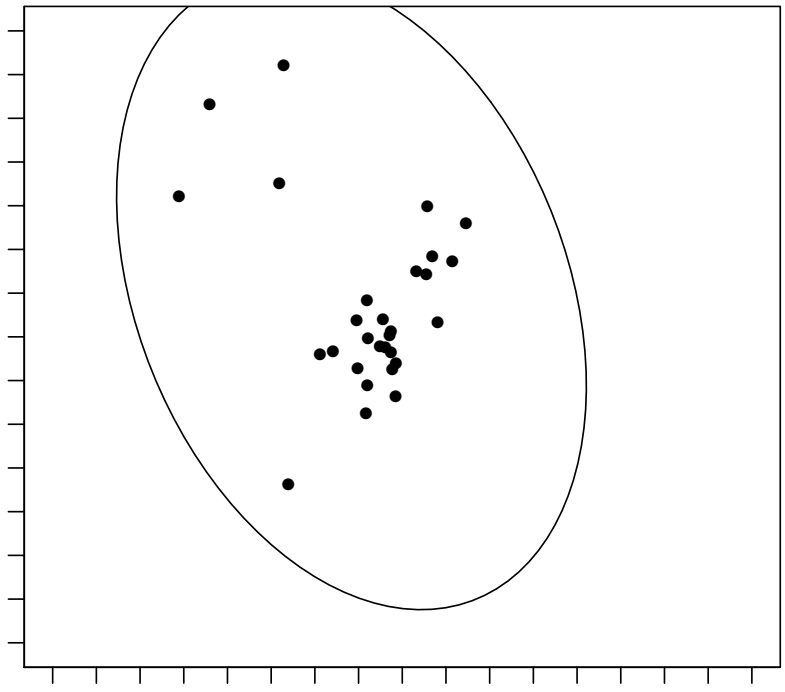
Regard again the two measurements during a test:



- Label $\boldsymbol{x}$ as outlier $D^{Mah}(\boldsymbol{x}|\boldsymbol{X}) < \min(D^{Mah})$.
- With less available data, *e.g.* at the beginning of the production process, when abnormal behavior is in addition more likely.

Detection of multivariate outliers: Mahalanobis depth

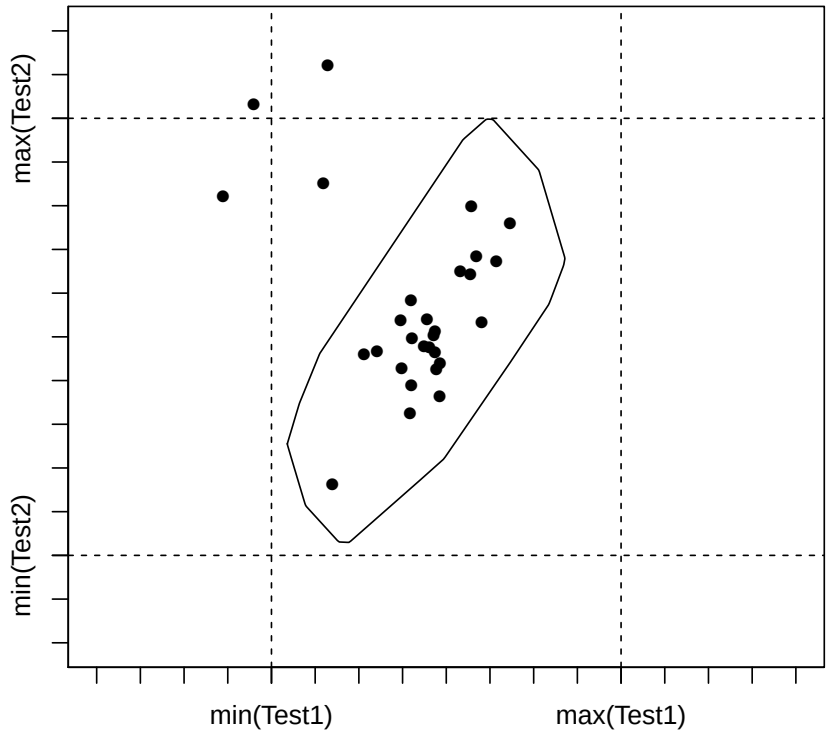
Regard again the two measurements during a test:



- Label $\boldsymbol{x}$ as outlier $D^{Mah}(\boldsymbol{x}|\boldsymbol{X}) < \min(D^{Mah})$.
- With less available data, *e.g.* at the beginning of the production process, when abnormal behavior is in addition more likely.
- Mahalanobis depth upper-level set is not robust.

Detection of multivariate outliers: projection depth

Regard again the two measurements during a test:

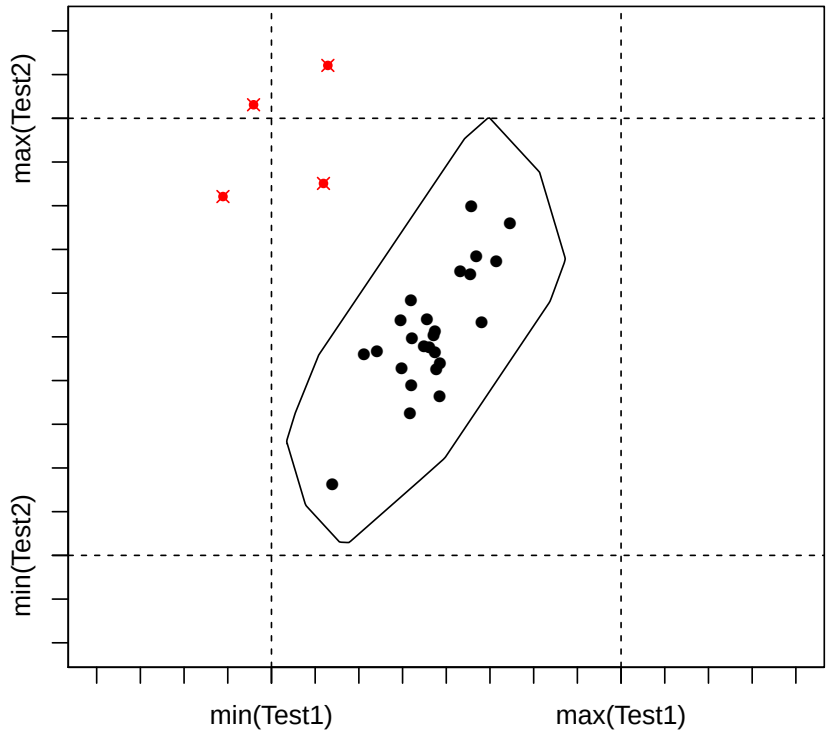


- Projection depth of $\boldsymbol{x}$ w.r.t. $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^n$:

$$D^{prj}(\boldsymbol{x}|\boldsymbol{X}) = \left(1 + \max_{\boldsymbol{u}\in\mathcal{S}^{d-1}} \frac{|\boldsymbol{x}^\top \boldsymbol{u} - \text{med}(\boldsymbol{X}\boldsymbol{u})|}{\text{MAD}(\boldsymbol{X}\boldsymbol{u})}\right)^{-1}.$$

Detection of multivariate outliers: projection depth

Regard again the two measurements during a test:



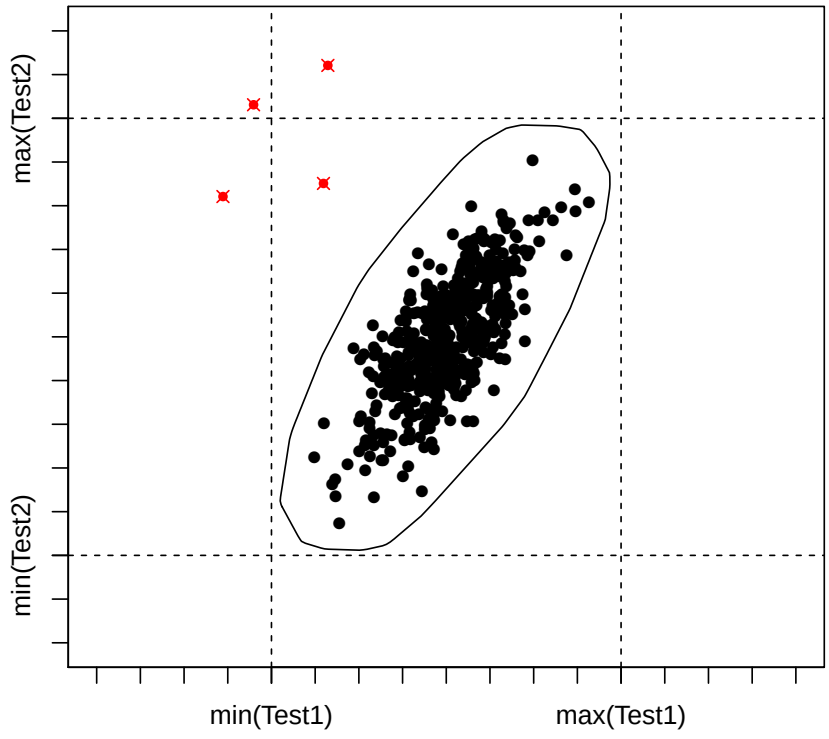
- Projection depth of $\boldsymbol{x}$ w.r.t. $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^n$:

$$D^{prj}(\boldsymbol{x}|\boldsymbol{X}) = \left(1 + \max_{\boldsymbol{u}\in\mathcal{S}^{d-1}} \frac{|\boldsymbol{x}^\top \boldsymbol{u} - \mathrm{med}(\boldsymbol{X}\boldsymbol{u})|}{\mathrm{MAD}(\boldsymbol{X}\boldsymbol{u})}\right)^{-1}.$$

- Label $\boldsymbol{x}$ as outlier $D^{prj}(\boldsymbol{x}|\boldsymbol{X}) < \min(D^{prj})$.

Detection of multivariate outliers: projection depth

Regard again the two measurements during a test:



- Projection depth of $\boldsymbol{x}$ w.r.t. $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^n$:

$$D^{prj}(\boldsymbol{x}|\boldsymbol{X}) = \left(1 + \max_{\boldsymbol{u}\in\mathcal{S}^{d-1}} \frac{|\boldsymbol{x}^\top \boldsymbol{u} - \mathrm{med}(\boldsymbol{X}\boldsymbol{u})|}{\mathrm{MAD}(\boldsymbol{X}\boldsymbol{u})}\right)^{-1}.$$

- Label $\boldsymbol{x}$ as outlier $D^{prj}(\boldsymbol{x}|\boldsymbol{X}) < \min(D^{prj})$.
- Stable rule for more observations.

Contents

The concept of data depth

Some examples
- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

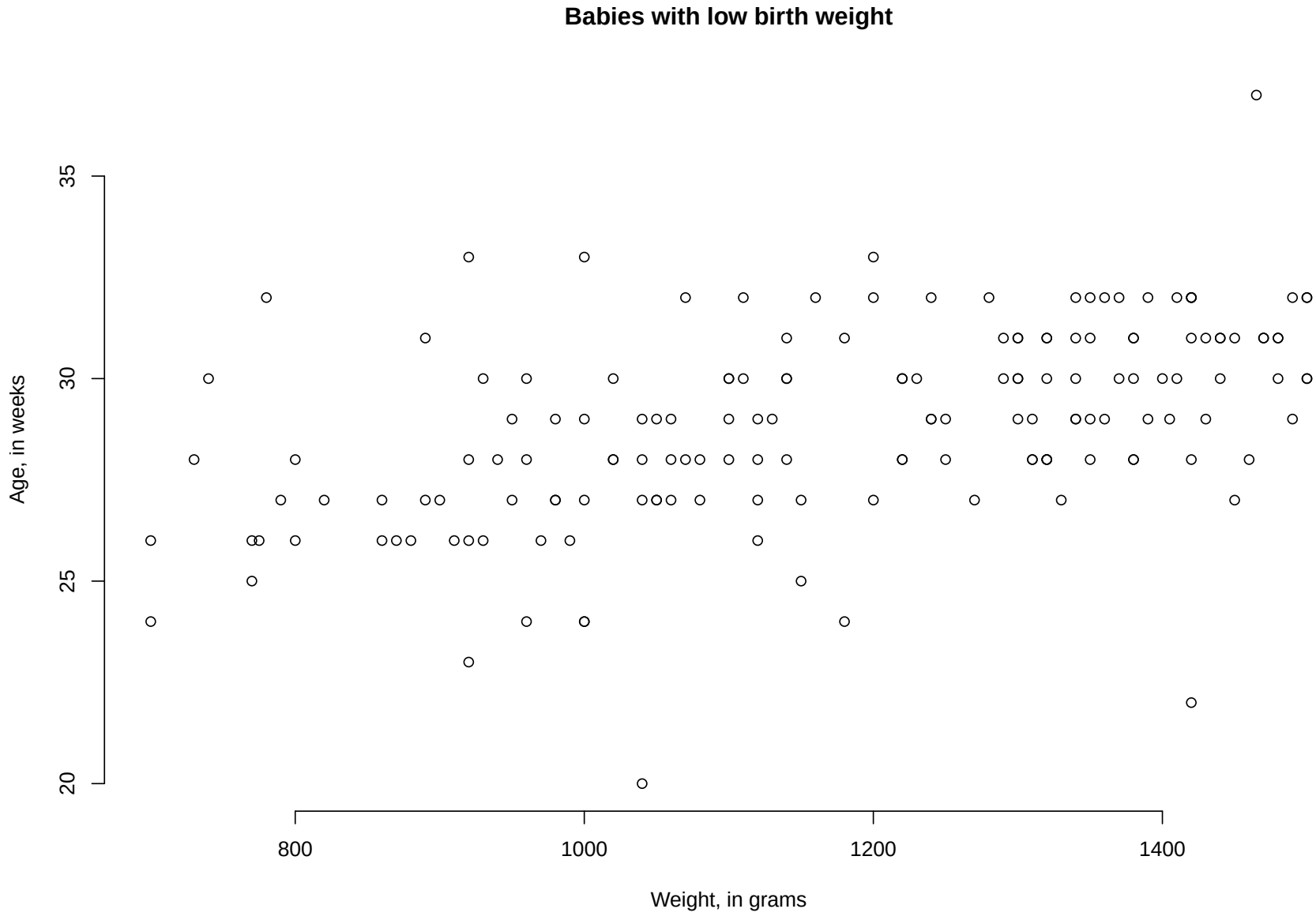
Halfspace depth
- Properties
- Trimmed regions

Classes of depth functions
- Depths satisfying the projection property

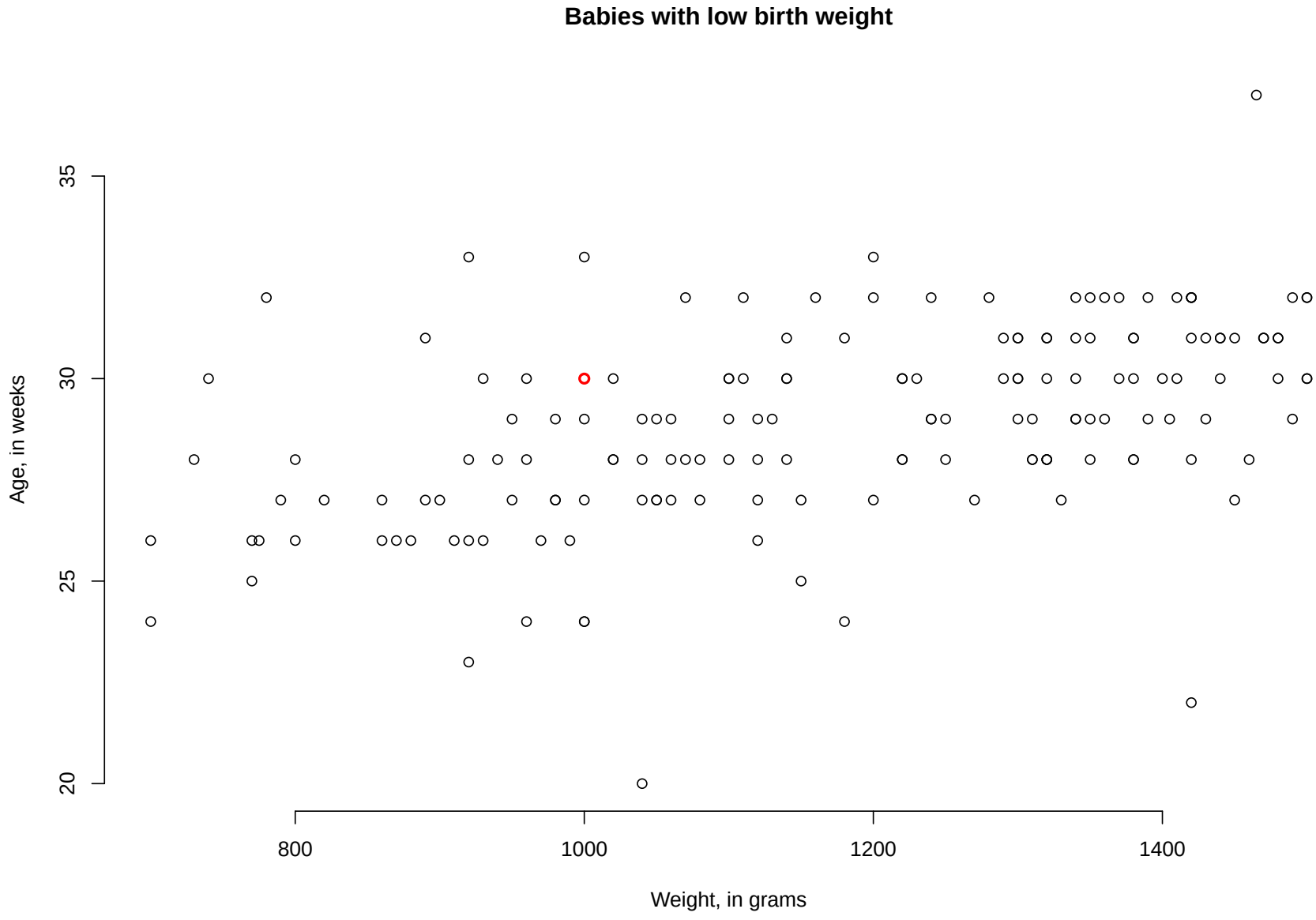
A concept of depth for the functional framework
- Functional outlier detection

References

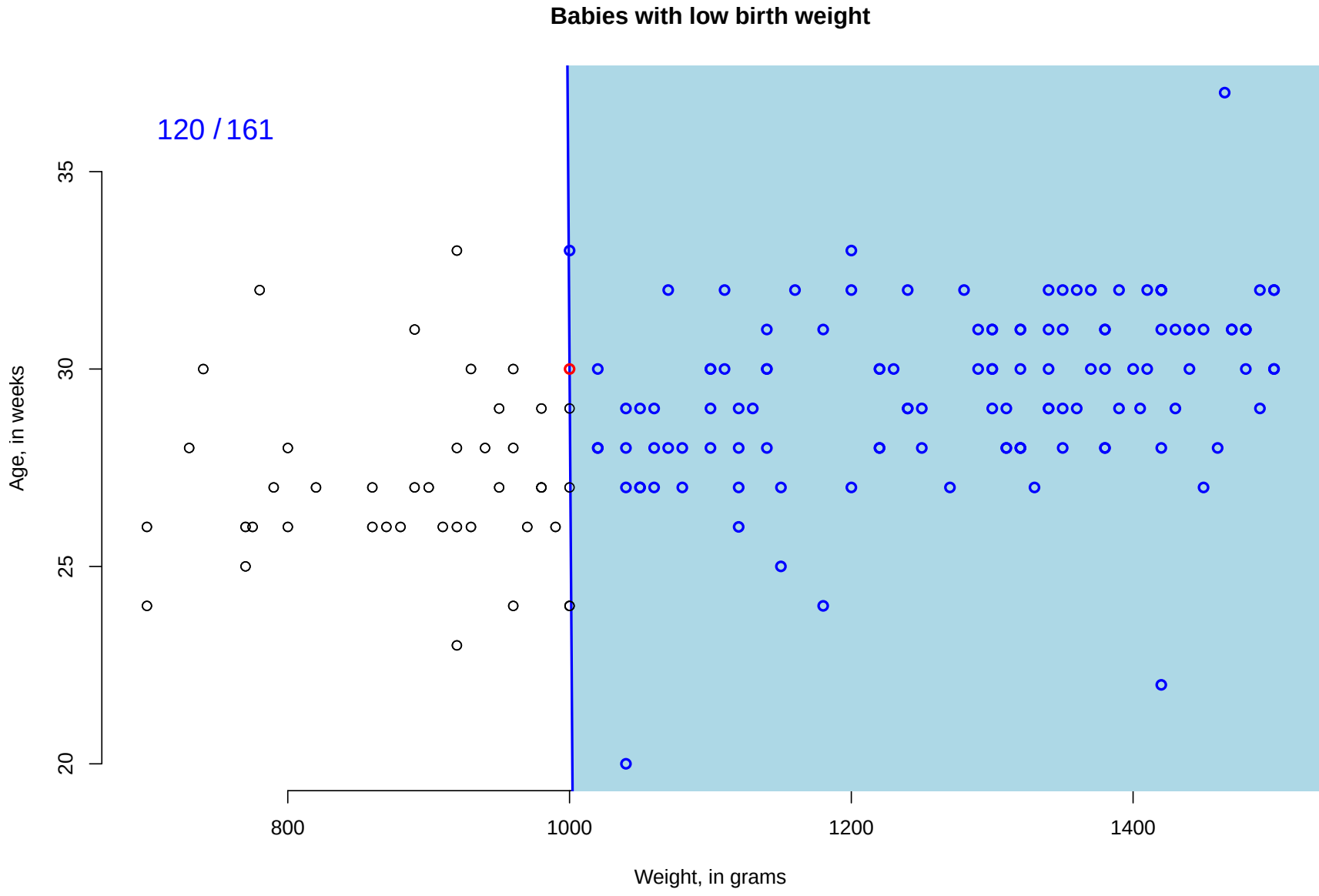
Tukey (=halfspace, location) data depth



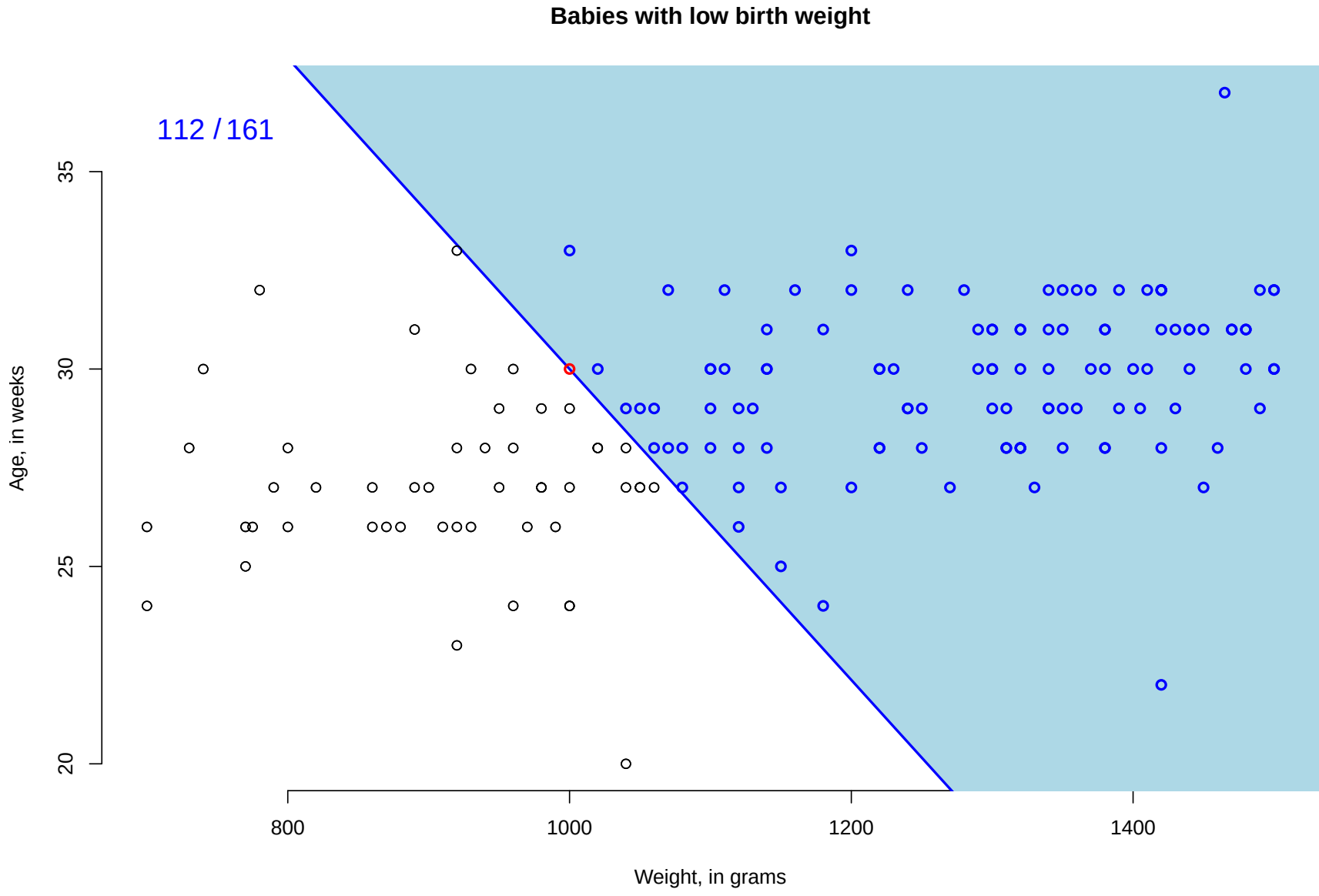
Tukey (=halfspace, location) data depth



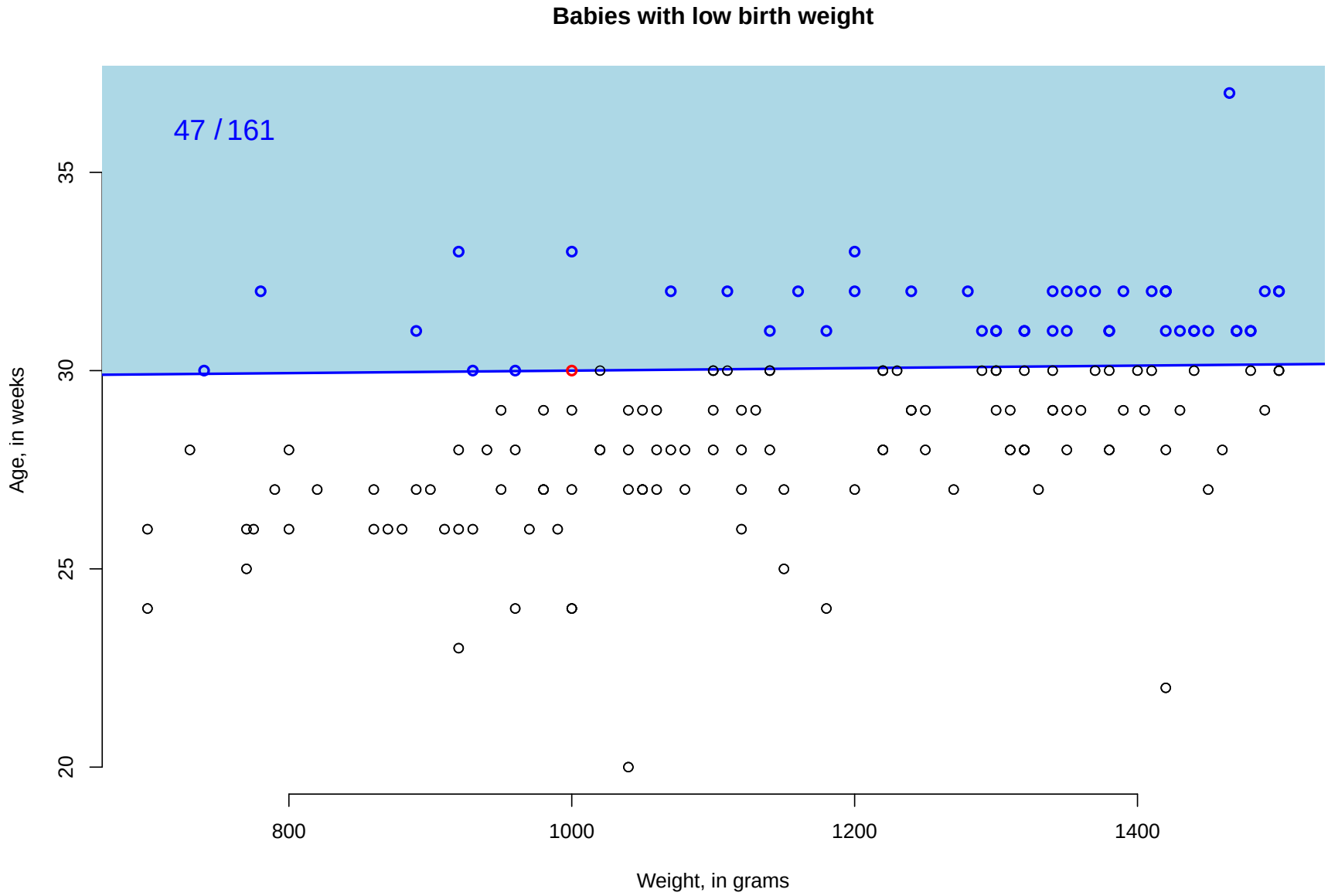
Tukey (=halfspace, location) data depth



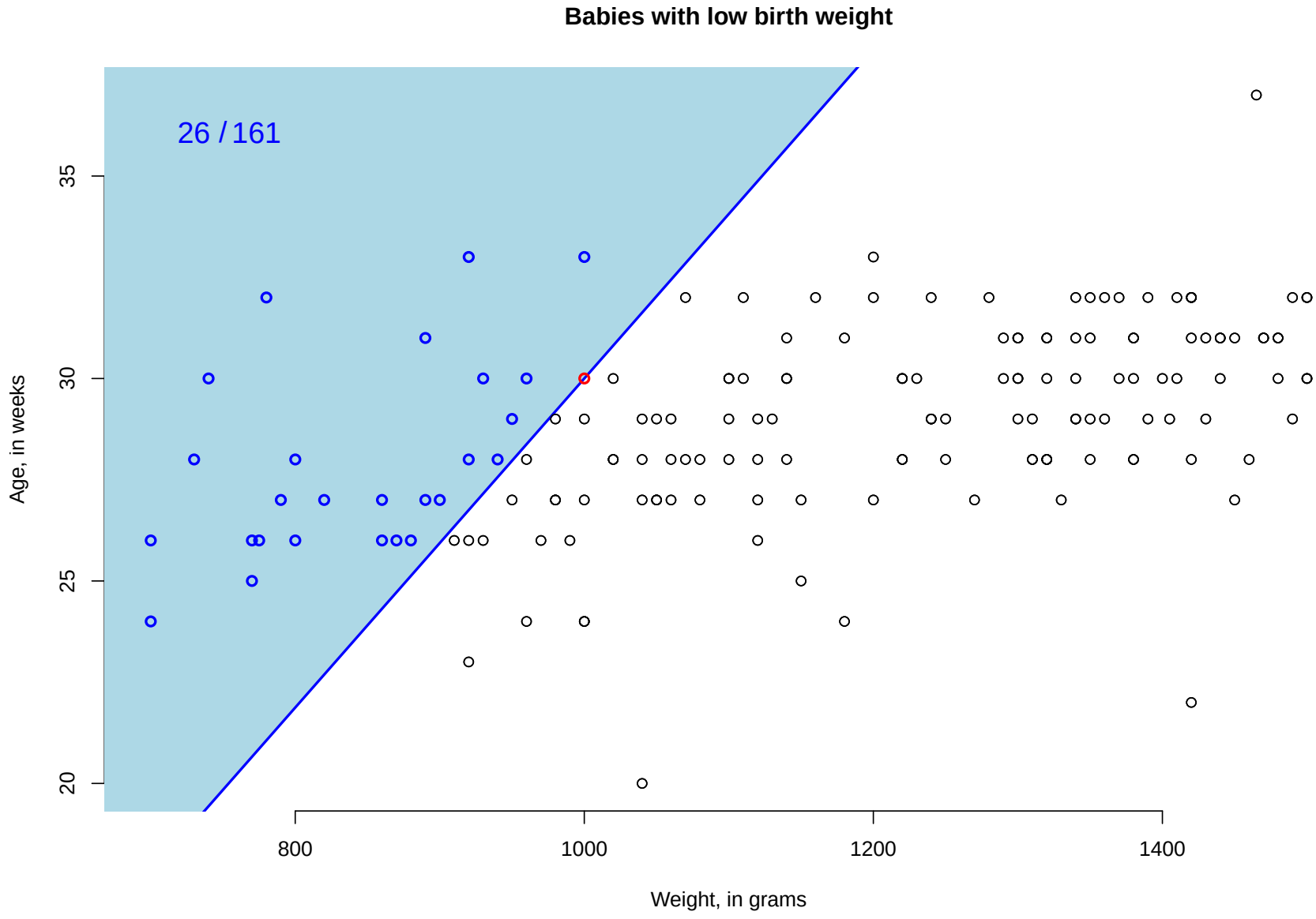
Tukey (=halfspace, location) data depth



Tukey (=halfspace, location) data depth



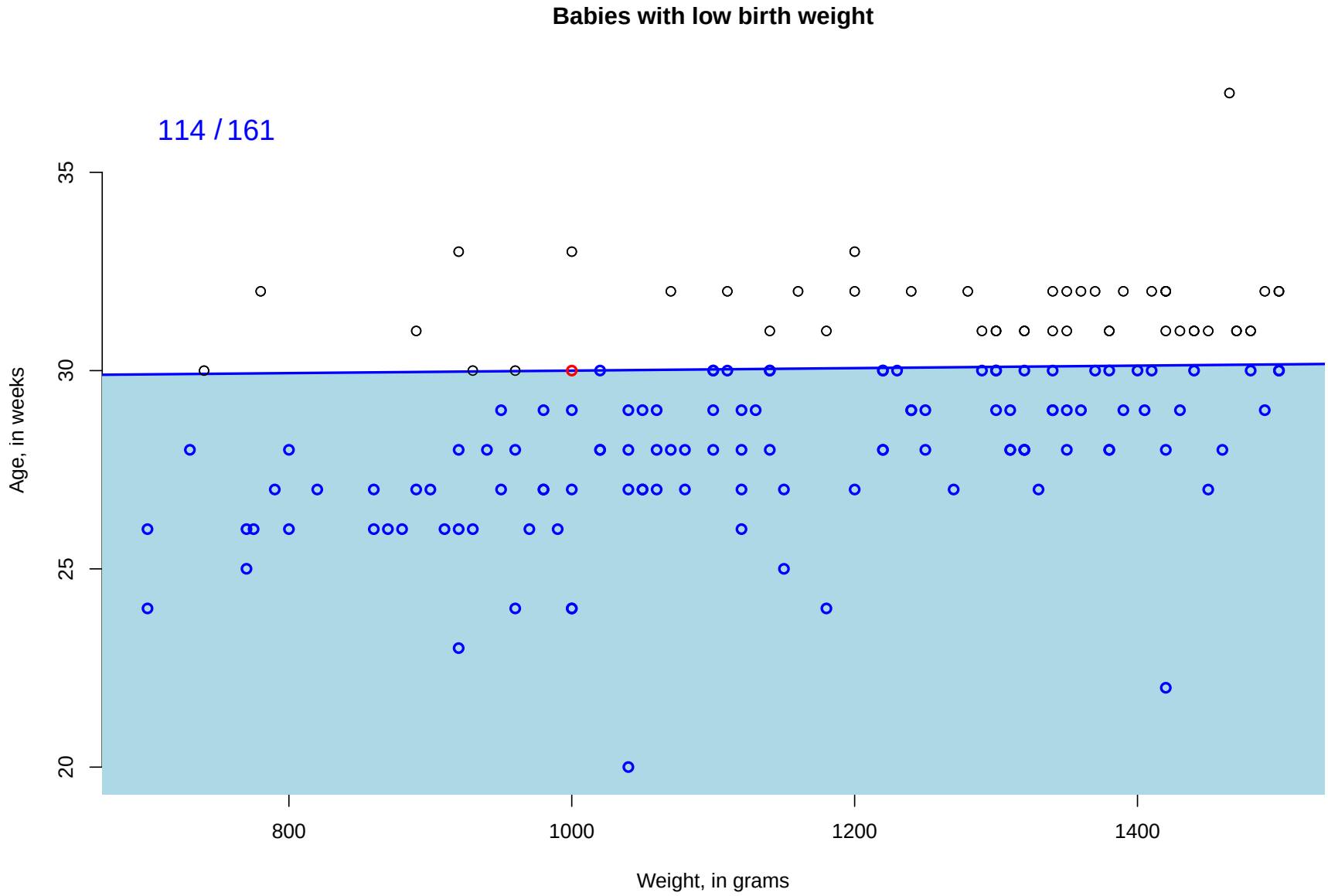
Tukey (=halfspace, location) data depth



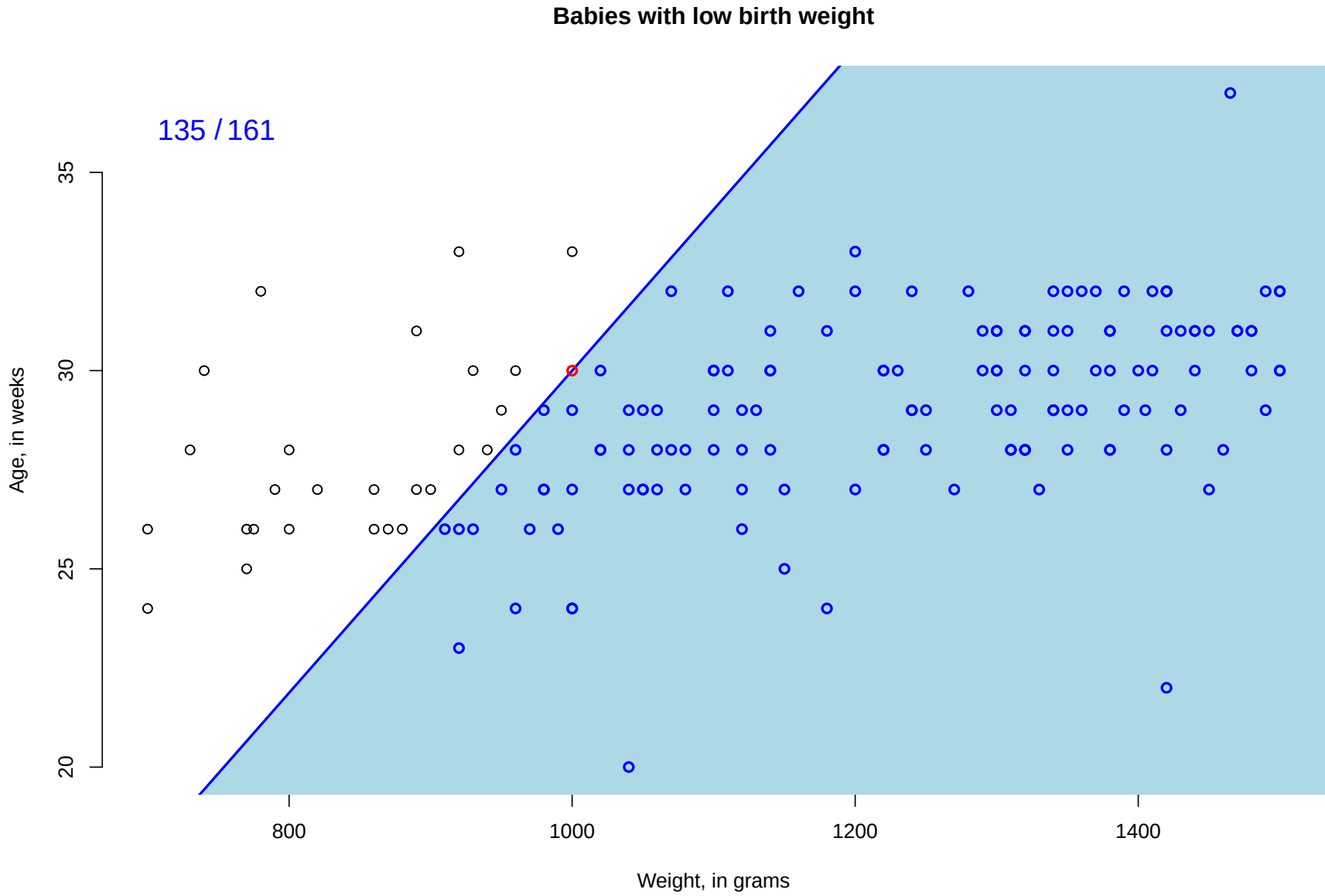
Tukey (=halfspace, location) data depth

Tukey (=halfspace, location) data depth

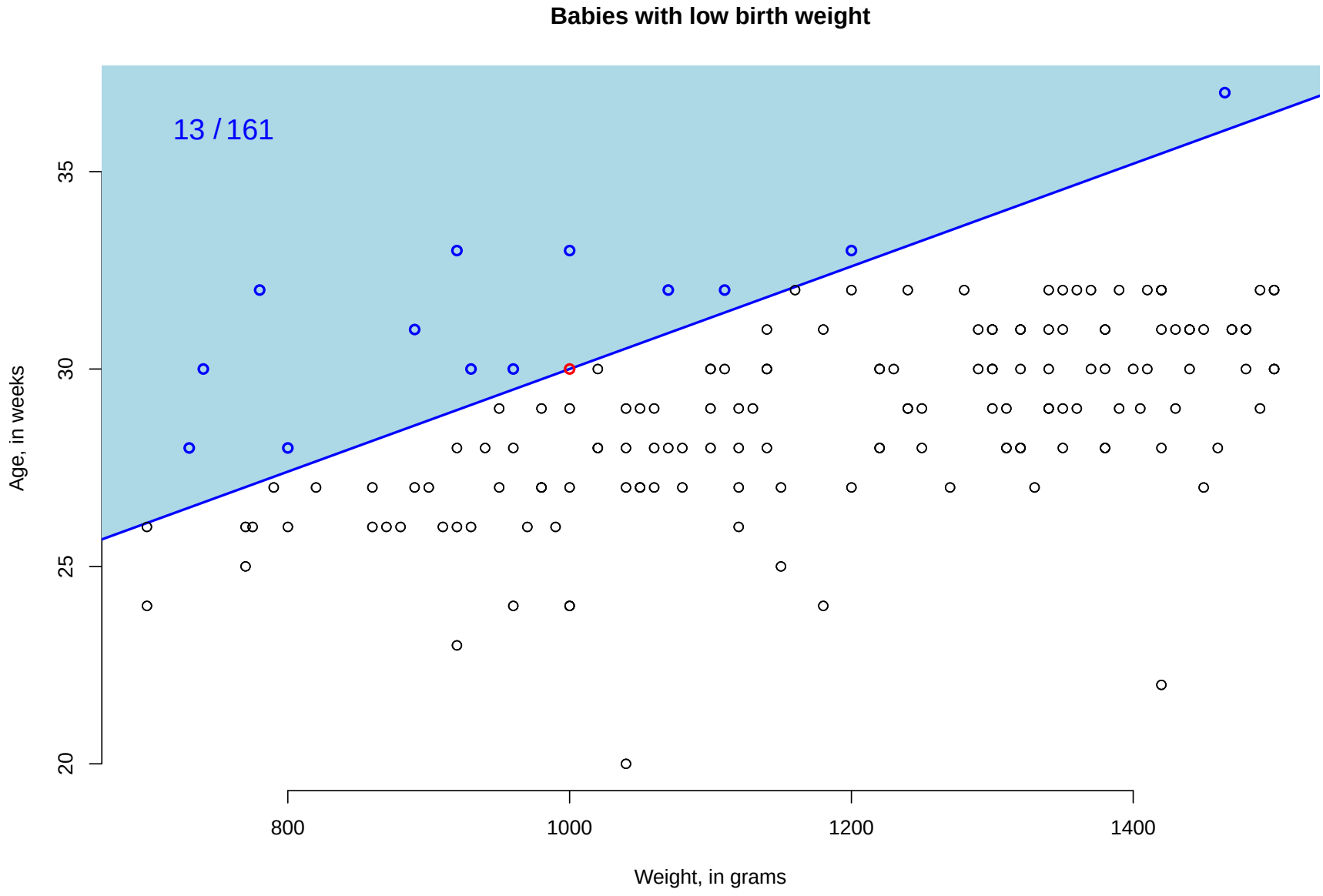
Tukey (=halfspace, location) data depth



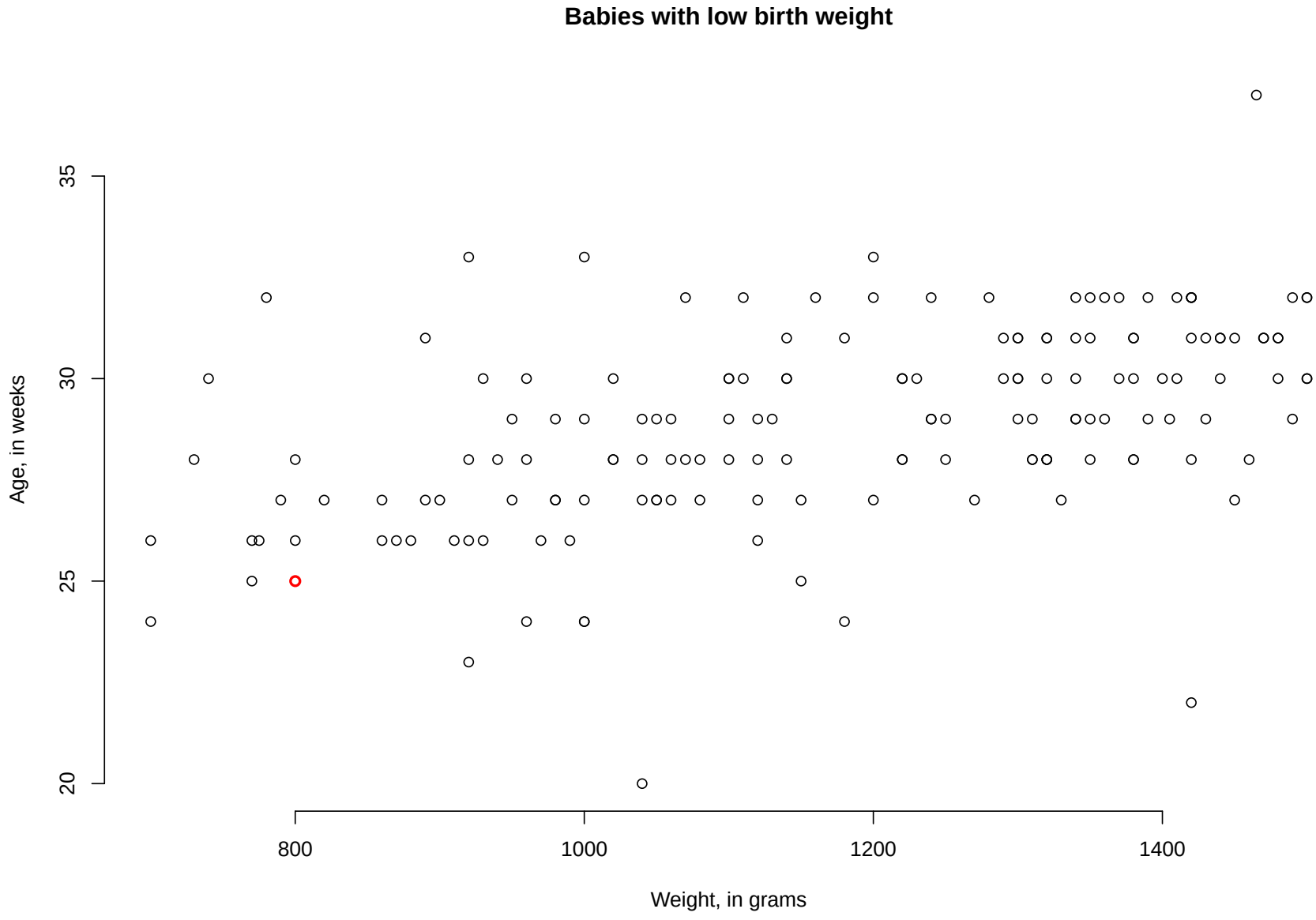
Tukey (=halfspace, location) data depth



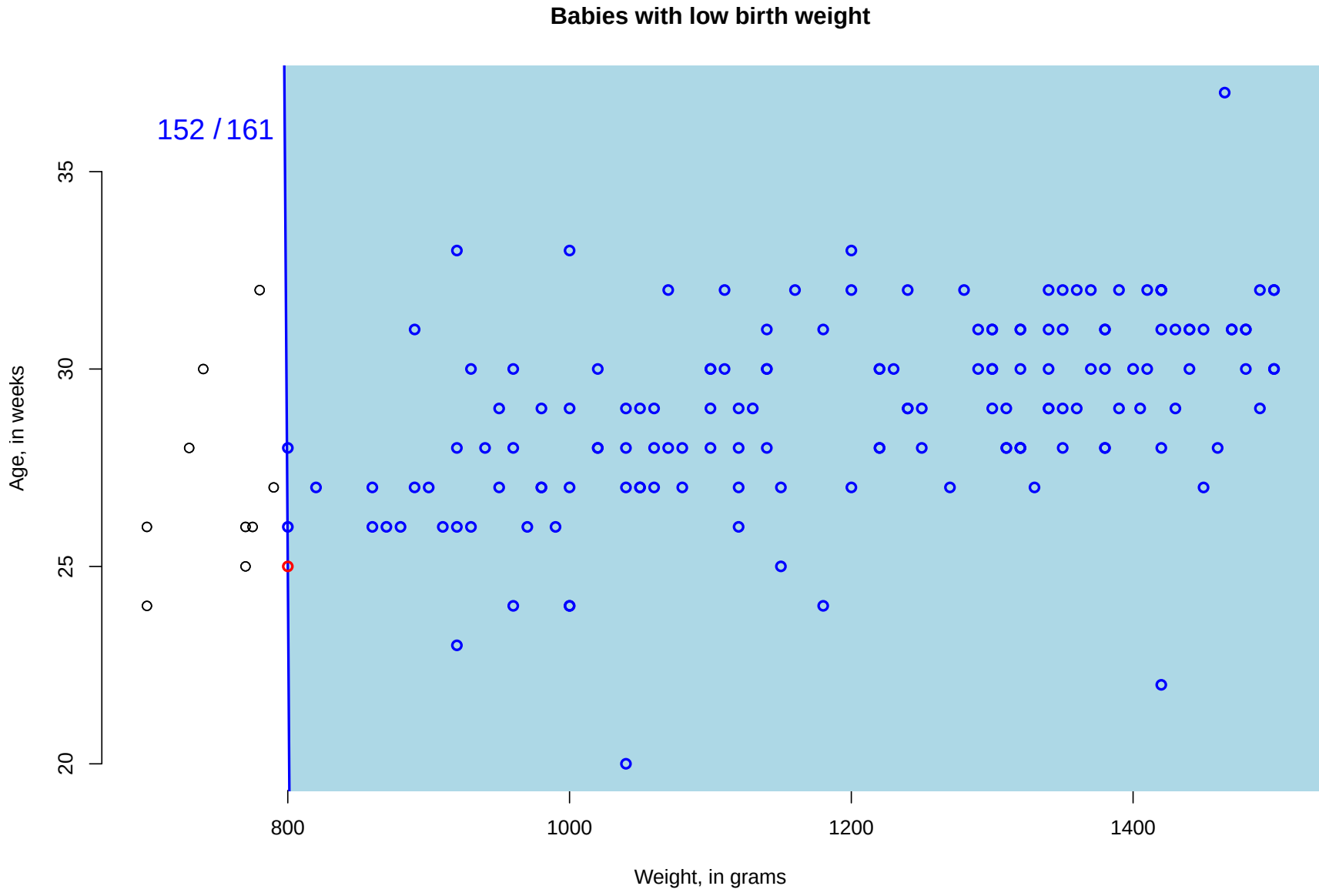
Tukey (=halfspace, location) data depth



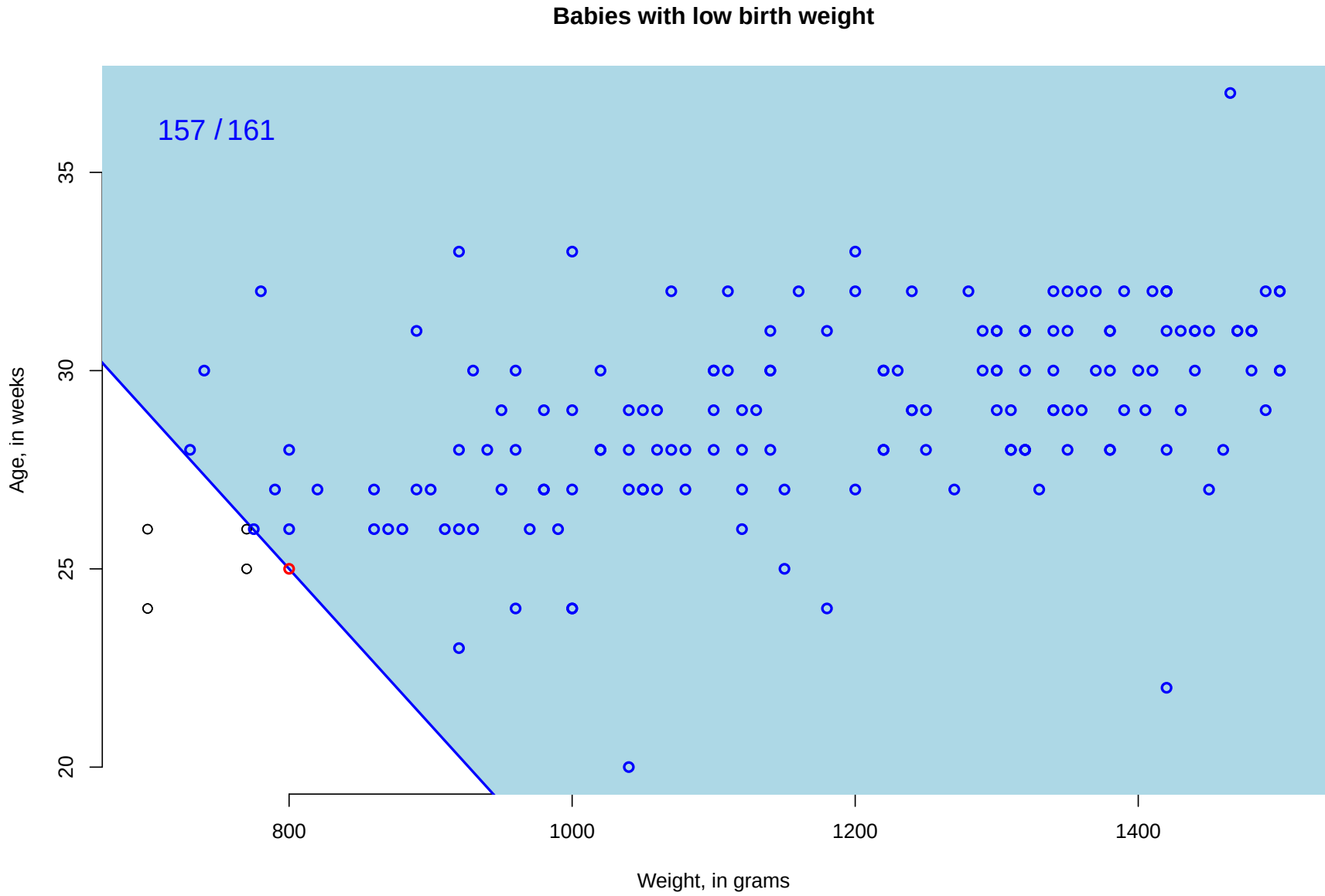
Tukey (=halfspace, location) data depth



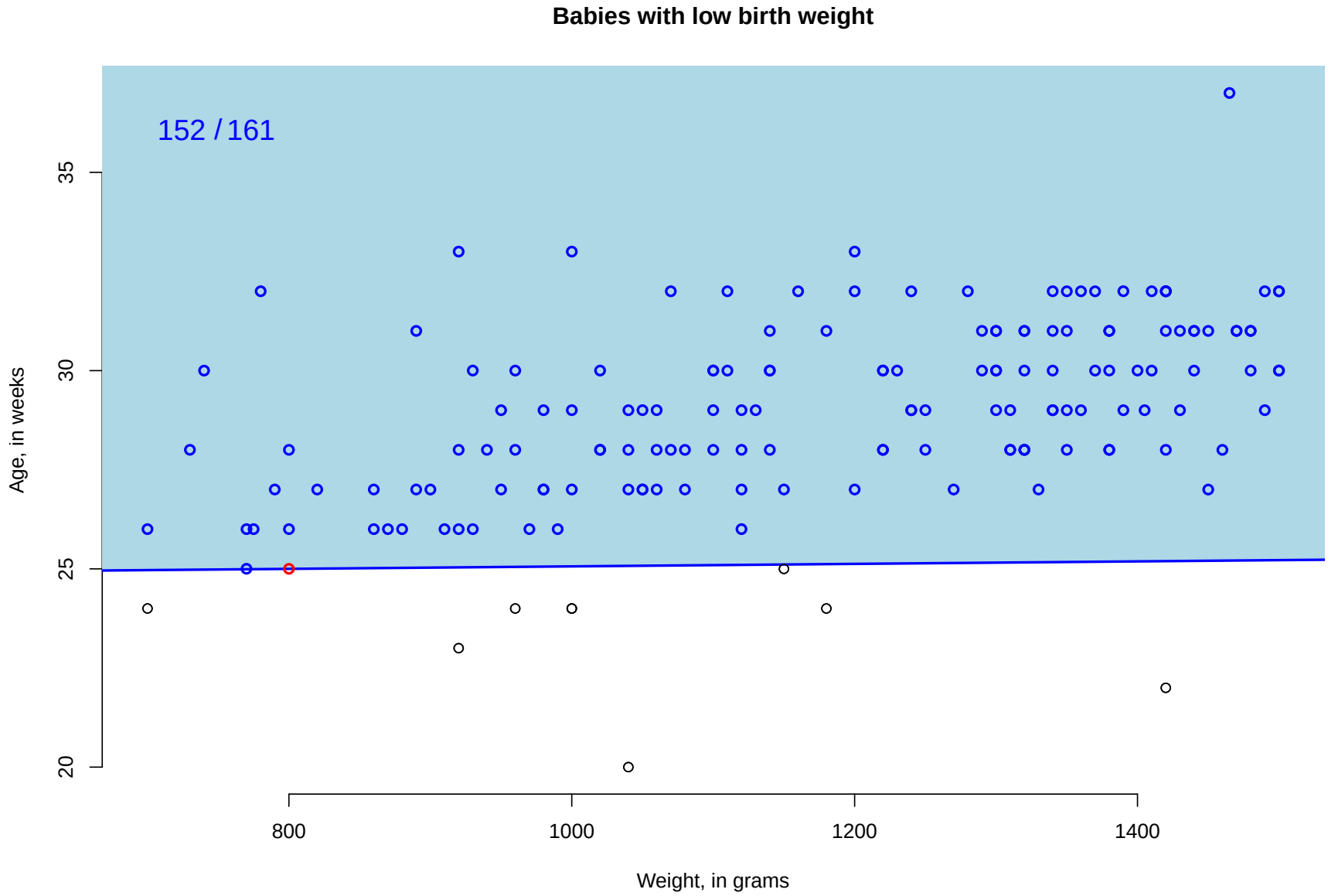
Tukey (=halfspace, location) data depth



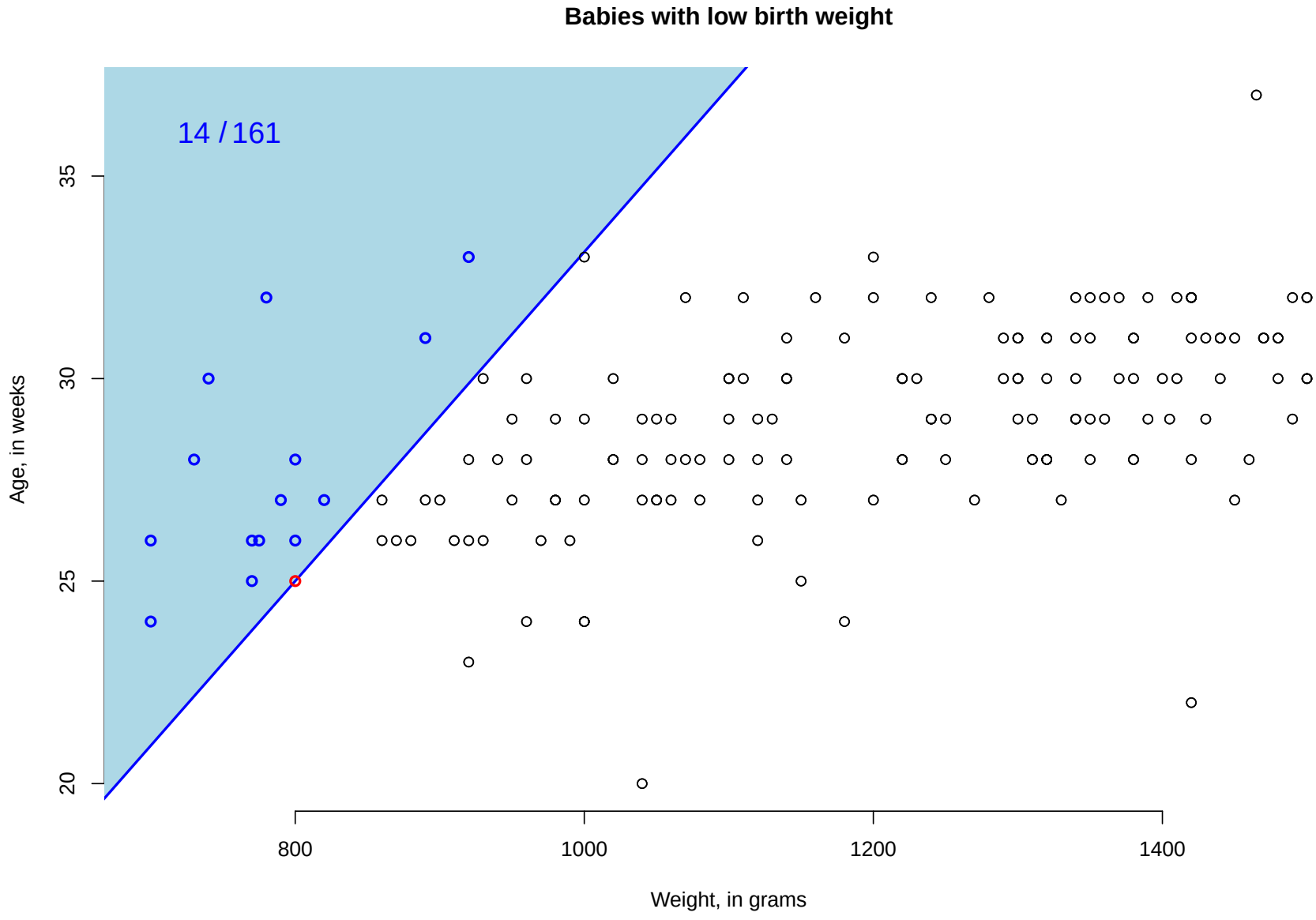
Tukey (=halfspace, location) data depth



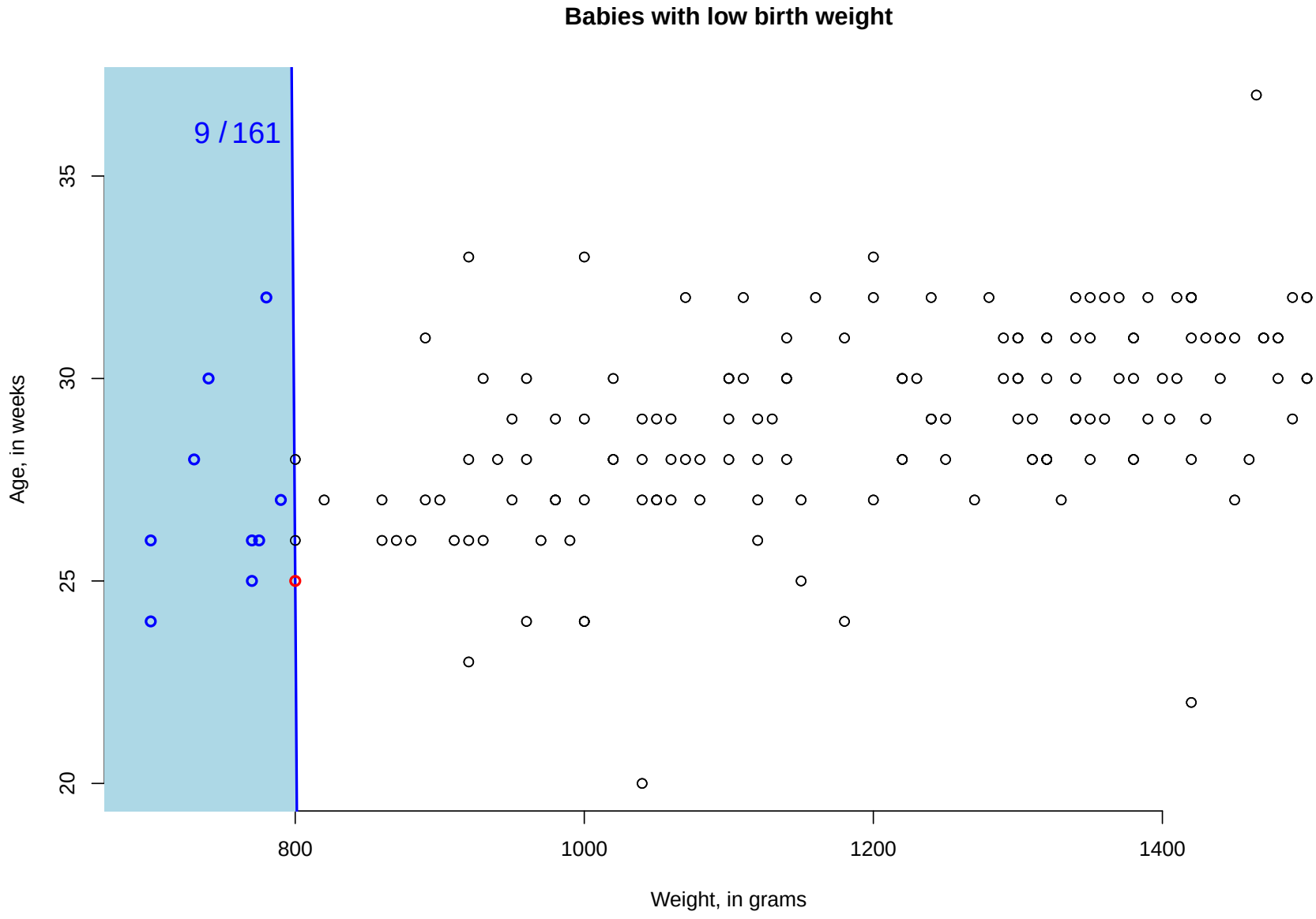
Tukey (=halfspace, location) data depth



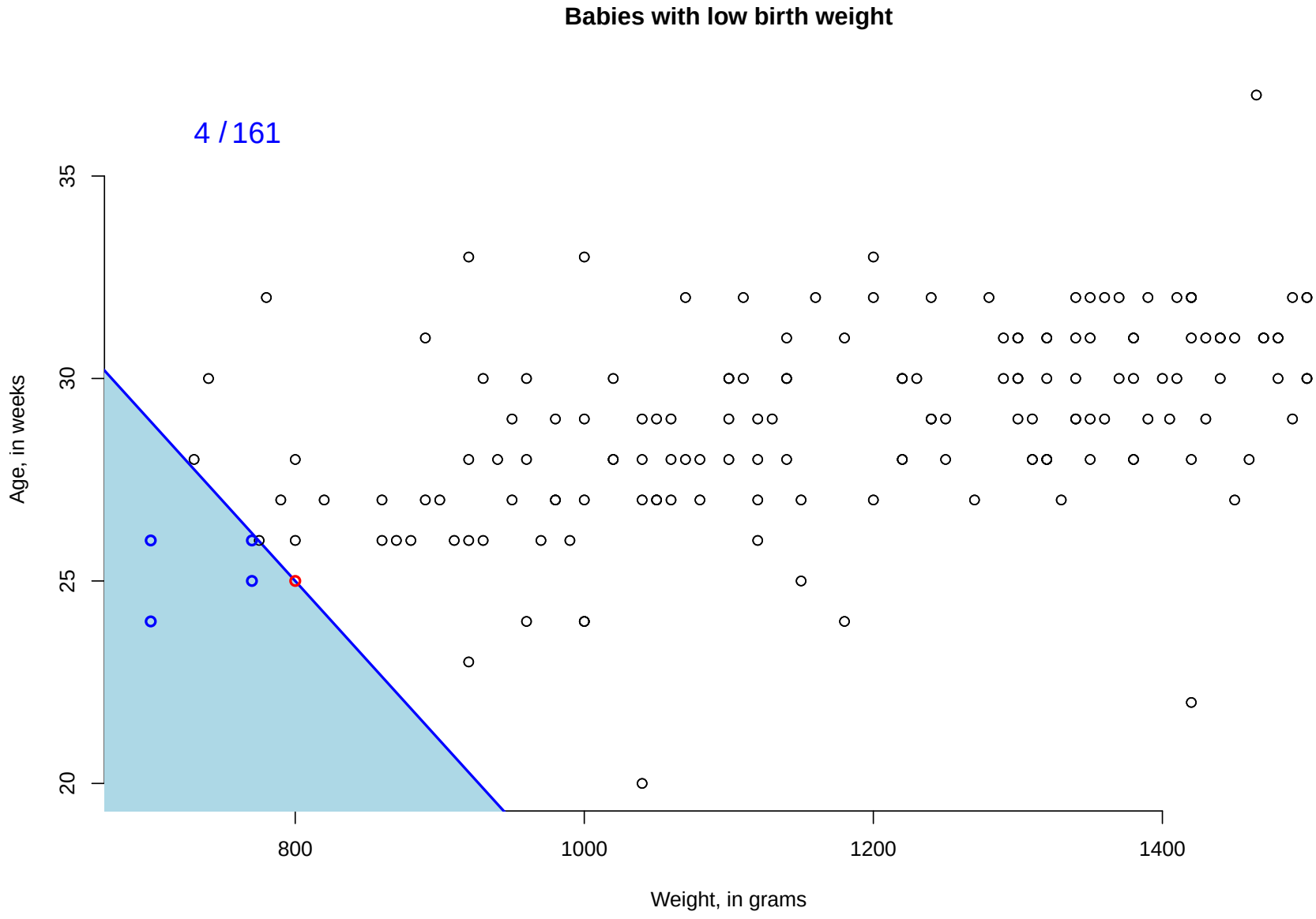
Tukey (=halfspace, location) data depth



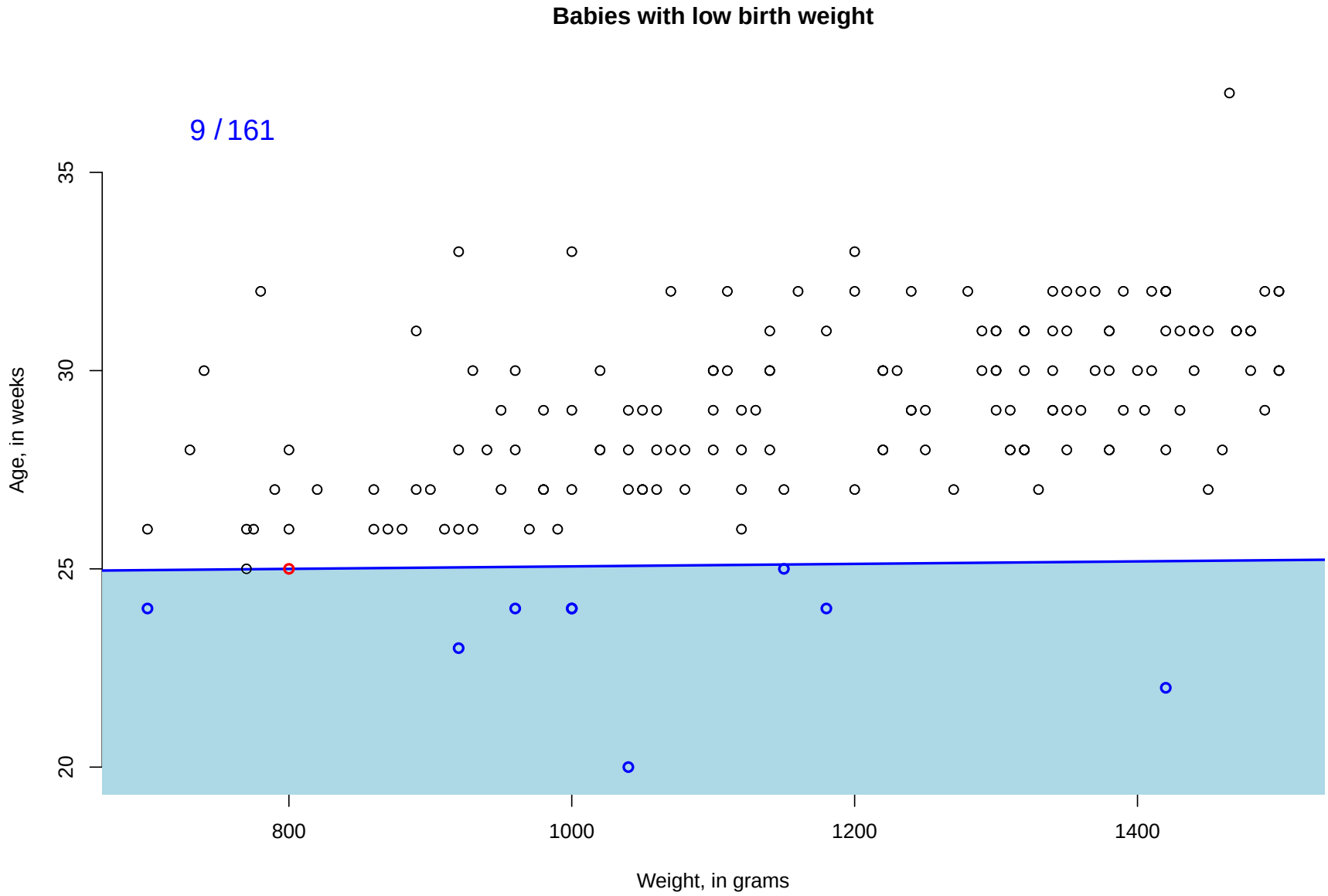
Tukey (=halfspace, location) data depth



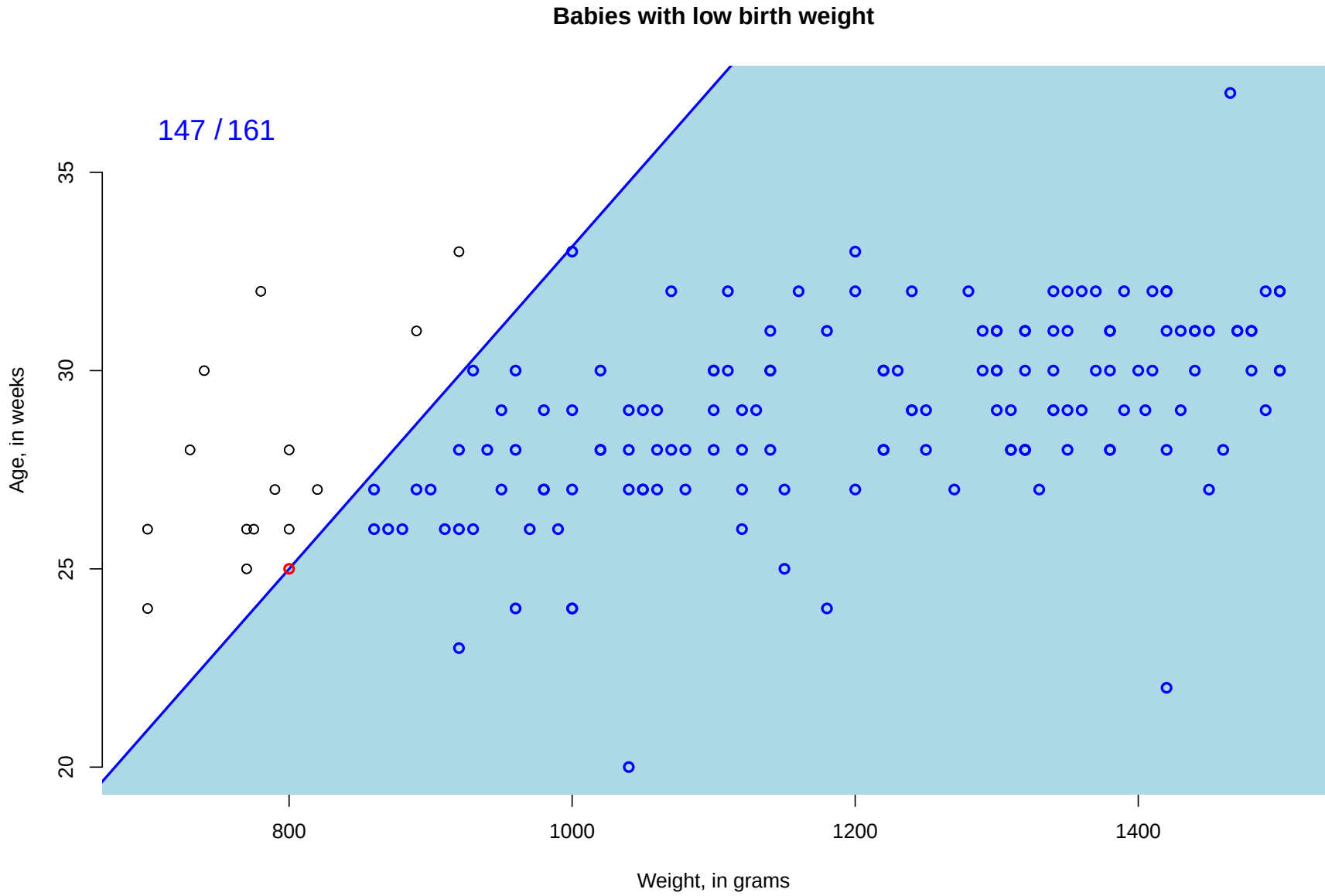
Tukey (=halfspace, location) data depth



Tukey (=halfspace, location) data depth

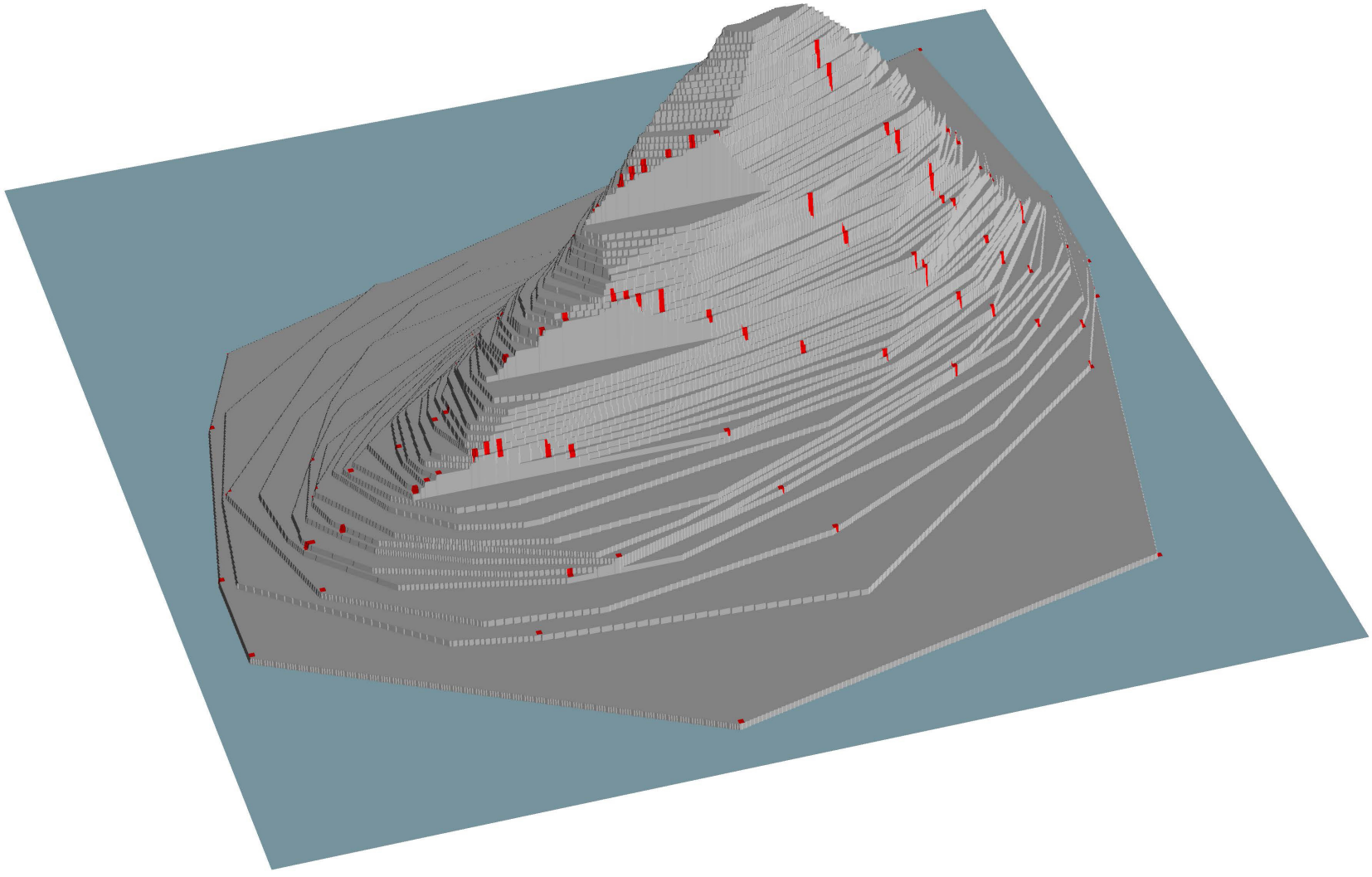


Tukey (=halfspace, location) data depth



Tukey (=halfspace, location) data depth

Tukey (=halfspace, location) data depth



Tukey (halfspace, location) depth

Tukey (1975) — "Mathematics and the picturing of data"

Tukey depth of $\boldsymbol{x} \in \mathbb{R}^d$ w.r.t. a d-variate random vector X distributed as P is defined as the smallest probability mass of a closed halfspace containing $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\},$$

and w.r.t. a sample $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\} \in \mathbb{R}^d$:

$$D^{Tuk(n)}(\boldsymbol{x}|\boldsymbol{X}) = \frac{1}{n} \min_{\boldsymbol{u} \in S^{d-1}} \sharp\{i : \boldsymbol{u}^T \boldsymbol{x}_i \geq \boldsymbol{u}^T \boldsymbol{x}\}.$$

Tukey depth defines a family of (depth-)trimmed (central) regions $D_\alpha^{Tuk}(X)$, the upper-level sets of the depth function:

$$D_\alpha^{Tuk}(X) = \left\{\boldsymbol{x} \in \mathbb{R}^d \,:\, D^{Tuk}(\boldsymbol{x}|X) \geq \alpha\right\}.$$

Contents

The concept of data depth

Some examples

- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth

- Properties
- Trimmed regions

Classes of depth functions

- Depths satisfying the projection property

A concept of depth for the functional framework

- Functional outlier detection

References

Halfspace depth: Properties

- satisfies D1 – D5 and D4con;
- $D^{Tuk}(\boldsymbol{x}|X) = 0$ for all $\boldsymbol{x} \notin \text{conv}\big(\text{supp}(X)\big)$;
- $D^{Tuk}(\cdot|X)$ attains its supremum (Rousseeuw & Ruts, 1999);
- at least one sample point has depth $\geq \frac{1}{d+1}$ (Mizera, 2002);
- if distribution smoothness condition is satisfied, i.e.

$$P(\partial H) = 0 \text{ for all } H \in \mathcal{H}\,, \mathcal{H} \text{ class of closed halfspaces in } \mathbb{R}^d\,,$$

  then the function $\boldsymbol{x} \mapsto D^{Tuk}(\boldsymbol{x}|\cdot)$ is continuous (Donoho & Gasko, 1992; Masse, 2004);
- determines the empirical distribution uniquely (Struyf & Rousseeuw, 1999; Koshevoy, 2002);
- is strongly uniformly consistent (Donoho & Gasko, 1992):

$$\sup_{\boldsymbol{x}\in\mathbb{R}^d} \left| D^{Tuk}(\boldsymbol{x}|P) - D^{Tuk}(\boldsymbol{x}|P_n) \right| \xrightarrow{a.s.} 0\,.$$

Uniform convergence rates (Burr & Fabrizio, 2017)

- Let $m(n)$ be the maximum number of subsets formed by intersecting finite sets of n points with halfspaces in $\mathbb{R}^d$.
  One has $m(n) \leq \frac{3n^{d+1}}{2(d+1)!}$.
- Let $\mathcal{H}$ be the set of all halfspaces in $\mathbb{R}^d$ and suppose $\epsilon > 0$. Then, for n sufficiently large,

$$P\Big(\sup_{H\in\mathcal{H}}\Big|\frac{1}{n}\#\{\boldsymbol{X}\cap H\} - P(X\in H)\Big| \geq \epsilon\Big) \leq 4m(n^2)e^{4\epsilon+4\epsilon^2}e^{-2n\epsilon^2}.$$

- X decays exponentially if $P\big(\|X\| > R\big) = O\big(e^{-\lambda\frac{R^2}{2}}\big)$.
- X is Lipschitz continuous in projection if $F_{X^T\boldsymbol{r}}$ is Lipschitz continuous $\forall\ \boldsymbol{r}\in S^{d-1}$.
- X is radially Lipschitz continuous if $F_{X^T\boldsymbol{r}}(t)$ is a Lipschitz continuous function of $\boldsymbol{r}$ for any fixed t.
- Suppose that X decays exponentially, is Lipschitz continuous in projection and radially. Fix ϵ, then there exists a constant C such that for n sufficiently large,

$$P\Big(\sup_{\boldsymbol{x}\in\mathbb{R}^d}\Big|D^{Tuk}(\boldsymbol{x}|X) - D^{Tuk}(\boldsymbol{x}|\boldsymbol{X})\Big| \geq \epsilon\Big) \leq Cn^{\frac{3}{2}(d-1)}e^{-2n\epsilon^2}.$$

Halfspace depth: Smoothness

- For a P on $\mathbb{R}^d$ and $\boldsymbol{x} \in \mathbb{R}^d$, the halfspace function of P at $\boldsymbol{x}$ is the function

$$\mathcal{G}(\boldsymbol{r}; \boldsymbol{x}, X) = P(X^T \boldsymbol{r} \geq \boldsymbol{x}^T \boldsymbol{r}) \quad \text{for} \quad \boldsymbol{r} \in S^{d-1}\,.$$

- Let X satisfy distribution smoothness condition, let $\arg\min_{\boldsymbol{r} \in S^{d-1}} \mathcal{G}(\boldsymbol{r}; \boldsymbol{x}, X)$ exist and be unique at some $\boldsymbol{x} \in \mathbb{R}$, and let $D^{Tuk}(\boldsymbol{x}|X) = \alpha > 0$, then (Masse, 2004)

$$\sqrt{n}\big(D^{Tuk}(\boldsymbol{x}|\boldsymbol{X}) - D^{Tuk}(\boldsymbol{x}|X)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \alpha(1-\alpha)\big)\,.$$

- P has contiguous support if there is no intersection of any two halfspaces with parallel boundaries that has nonempty interior but zero probability and divides the support of P into two parts.
- Halfspace depth characterizes P, if P is contiguous and all depth contours are smooth (Kong & Zuo, 2010).

Halfspace depth: Smoothness

- Let P be a contiguous distribution on $\mathbb{R}^d$, let $\boldsymbol{x} \in \mathbb{R}^d$ such that $D^{Tuk}(\boldsymbol{x}|X) = \alpha > 0$, and let $D^{Tuk}_{\alpha}(X)$ have non-empty interior. Then $\mathcal{G}(\boldsymbol{r}; \boldsymbol{x}, X)$ attains more then one different global minima at $\boldsymbol{x}$, *i.e.* for some $\boldsymbol{r}_1 \neq \boldsymbol{r}_2$

$$\liminf_{\boldsymbol{r} \to \boldsymbol{r}_1} \mathcal{G}(\boldsymbol{r}; \boldsymbol{x}, X) = \liminf_{\boldsymbol{r} \to \boldsymbol{r}_2} \mathcal{G}(\boldsymbol{r}; \boldsymbol{x}, X) = D^{Tuk}(\boldsymbol{x}|X)\,,$$

  if and only if the halfspace depth α-contour is not smooth at $\boldsymbol{x}$ (Gijbels & Nagy, 2016).
- Regard the condition
  "Halfspace depth contours of P are smooth for all $\boldsymbol{x} \in \mathbb{R}^d$".
- This condition (Gijbels & Nagy, 2016):
  - holds for elliptically symmetric densities;
  - does not hold for normal mixtures;
  - does not hold for L_p-symmetric distributions;
  - does not hold for centrally symmetric distribution with smooth quasi-concave density;
  - does not hold for strictly quasi-concave distributions.

Connection to the densest hemisphere problem

- For $\boldsymbol{x}$, $\boldsymbol{X}$, denote the set of directions attaining $D^{Tuk}(\boldsymbol{x}|\boldsymbol{X})$

$$R(\boldsymbol{x}, \boldsymbol{X}) = \underset{\boldsymbol{r} \in S^{d-1}}{\arg\min} \#\{i \,:\, \boldsymbol{x}_i^T \boldsymbol{r} \geq \boldsymbol{x}^T \boldsymbol{r}\}\,.$$

- The problem of computing $D^{Tuk}(\boldsymbol{x}|\boldsymbol{X})$ can be seen as invariant in more than just affine way.
- For $\boldsymbol{X}$ with $\boldsymbol{x}_i \neq \boldsymbol{x}$ for $i = 1, ..., n$ define:

$$\boldsymbol{Y} = \{\boldsymbol{y}_i \,|\, \boldsymbol{y}_i = \frac{\boldsymbol{x}_i - \boldsymbol{x}}{\|\boldsymbol{x}_i - \boldsymbol{x}\|}\}\,.$$

- One can easily show that

$$D^{Tuk}(\boldsymbol{x}|\boldsymbol{X}) = D^{Tuk}(\boldsymbol{0}|\boldsymbol{Y})\,.$$

- Calculating $D^{Tuk}(\boldsymbol{0}|\boldsymbol{Y})$ and picking $\boldsymbol{r} \in R(\boldsymbol{0}|\boldsymbol{Y})$ corresponds to finding the open densest hemisphere, *i.e.* an open hemisphere containing the highest portion of observations. It is of non-polynomial complexity (Johnson & Preparata, 1987).

Connection to the supervised classification

- Let $\boldsymbol{X}_1 = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_{n_1}\}$ and $\boldsymbol{X}_2 = \{\boldsymbol{x}_{n_1+1}, ..., \boldsymbol{x}_{n_1+n_2}\}$ be two training classes in $\mathbb{R}^d$.
- Regarding the univariate case, if the two classes are well separated, one would expect that most of the observed differences $\boldsymbol{x}_i - \boldsymbol{x}_j$, $i = 1, ..., n_1$, $j = n_1 + 1, ..., n_1 + n_2$ will have the same sign (positive or negative).
- Generalizing this idea to multivariate case, we define the separating hyperplane by $\boldsymbol{r} \in S^{d-1}$ and search $\boldsymbol{r}$ by maximizing (Ghosh & Chaudhuri, 2005):

$$U_{(n_1,n_2)}(\boldsymbol{r}) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_1+n_2} I\left(\boldsymbol{r}^T(\boldsymbol{x}_i - \boldsymbol{x}_j) > 0\right).$$

- Letting $\boldsymbol{Y} = \{\boldsymbol{x}_i - \boldsymbol{x}_j \,:\, i = 1, ..., n_1, j = n_1 + 1, ..., n_1 + n_2\}$ this is equivalent to calculating

$$D^{Tuk}(\boldsymbol{0}|\boldsymbol{Y})$$

and picking $\boldsymbol{r} \in R(\boldsymbol{0}|\boldsymbol{Y})$.

Contents

The concept of data depth

Some examples

- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth

- Properties
- Trimmed regions

Classes of depth functions

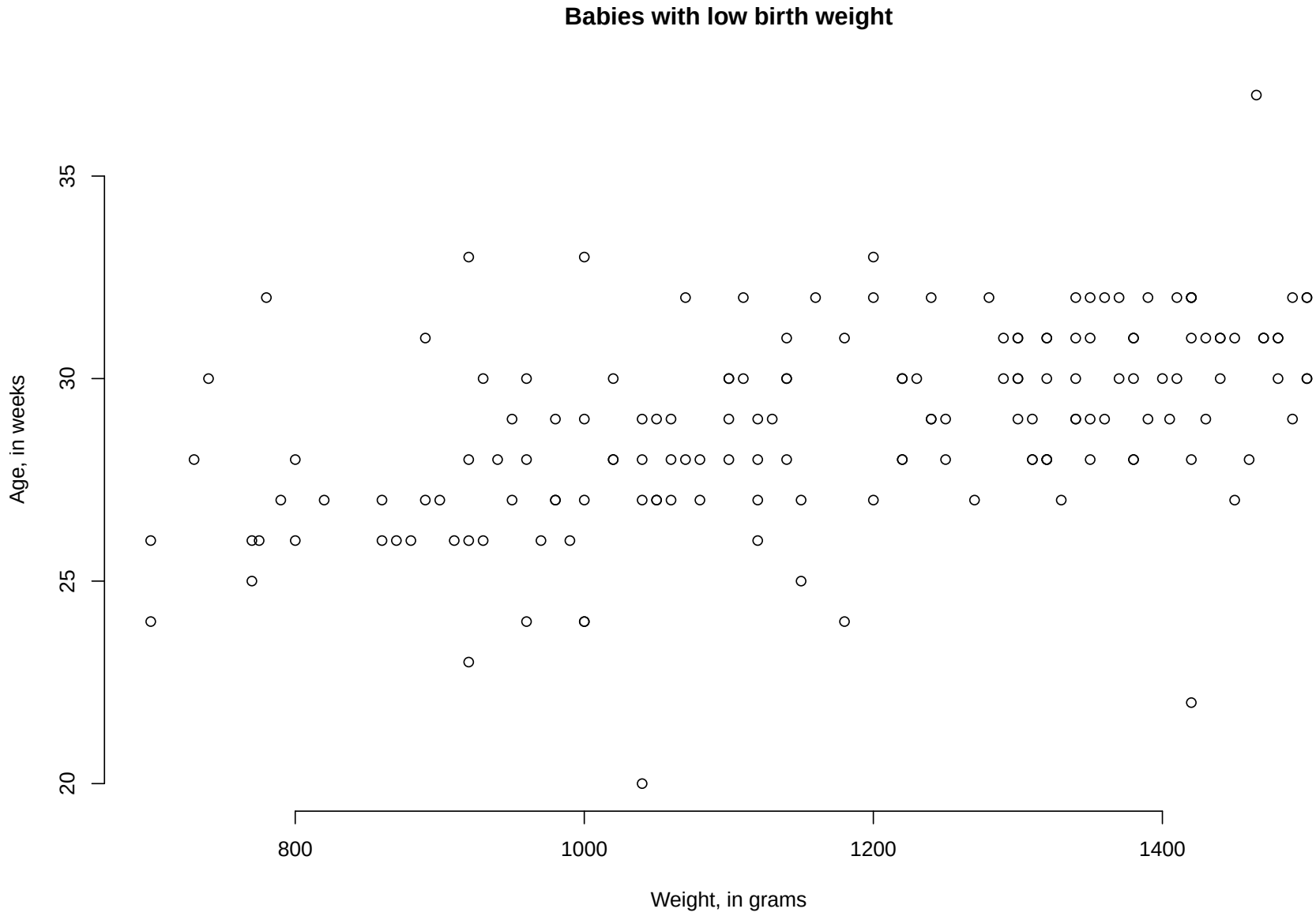
- Depths satisfying the projection property

A concept of depth for the functional framework

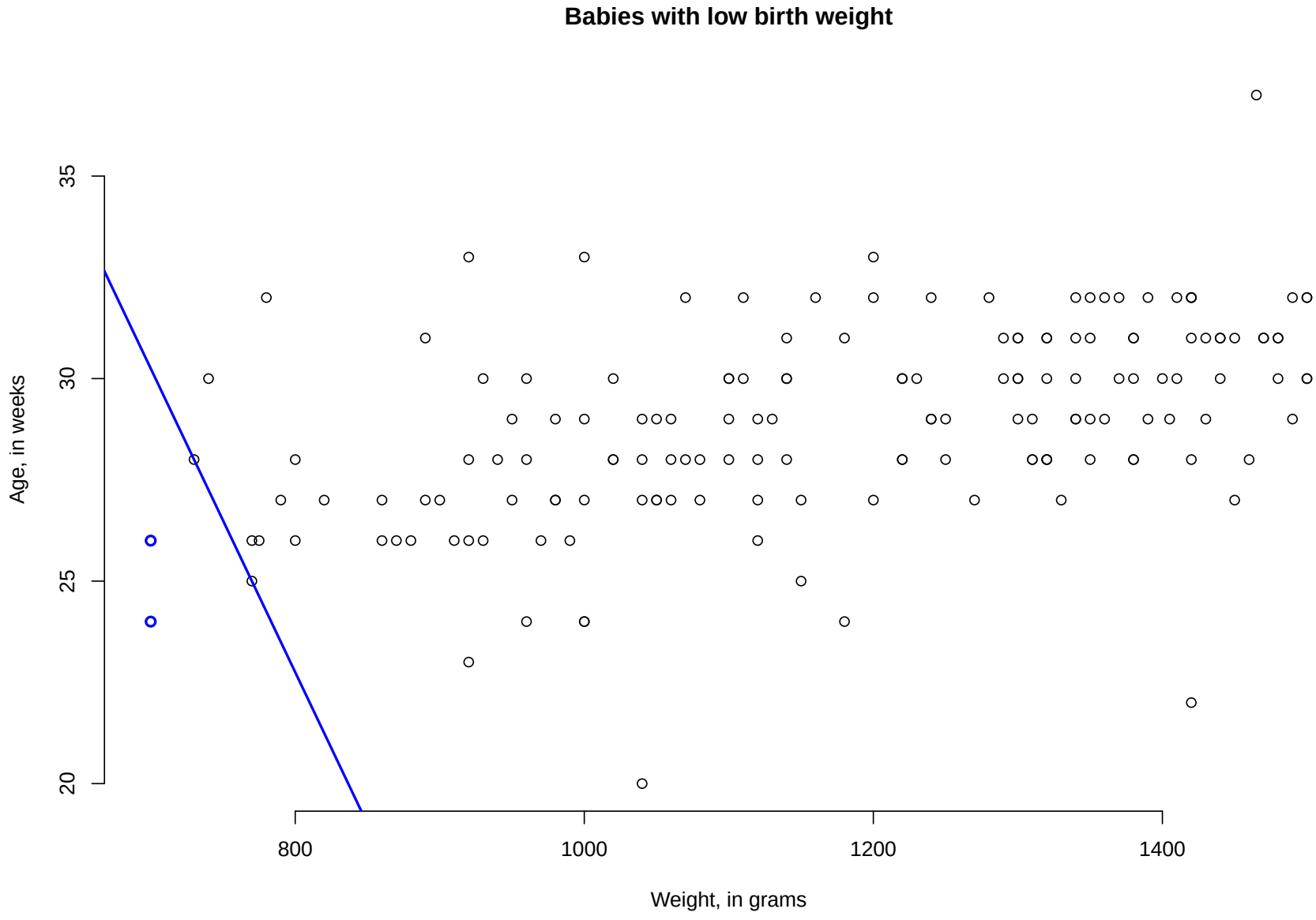
- Functional outlier detection

References

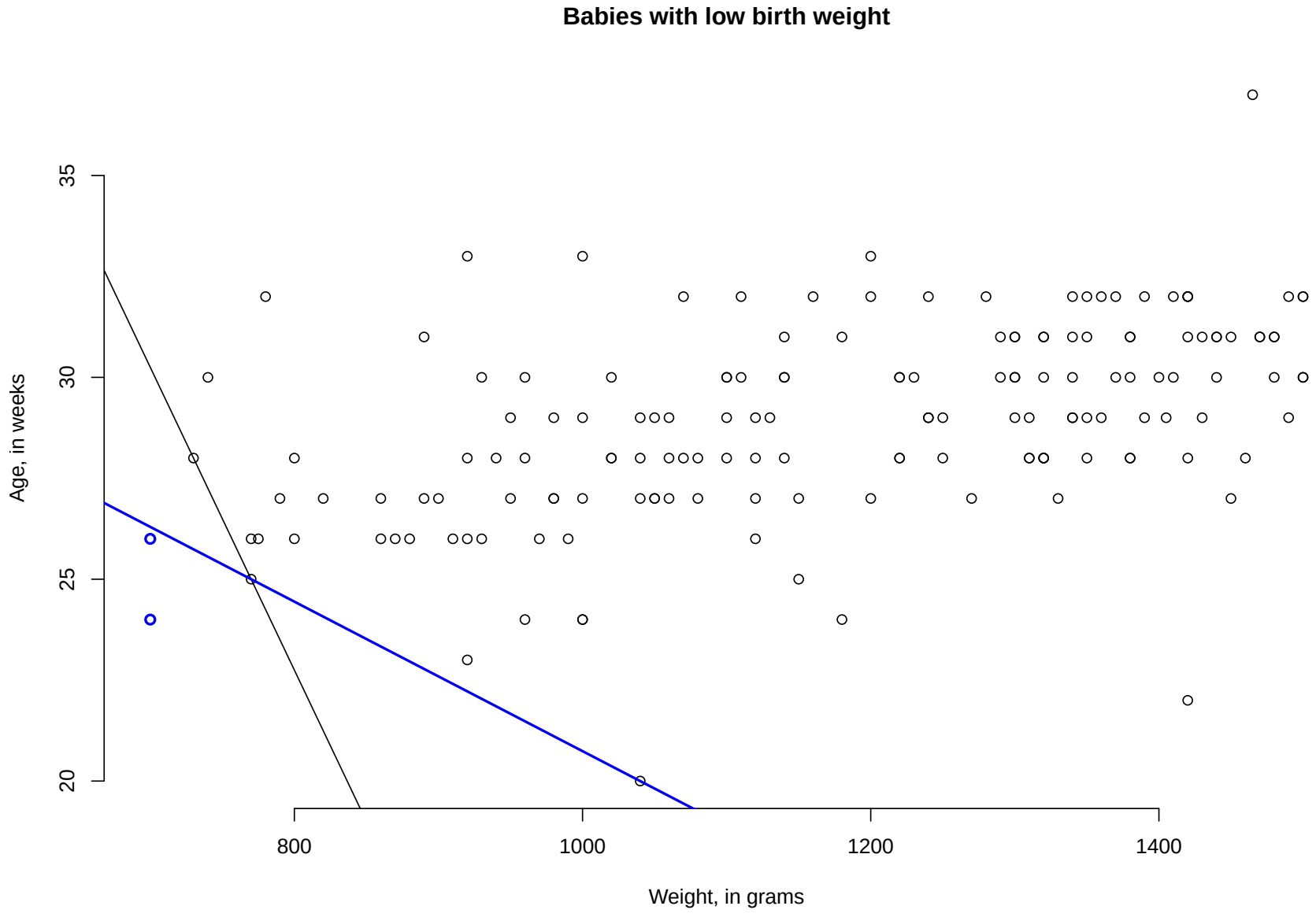
Tukey (=halfspace, location) depth-trimmed regions



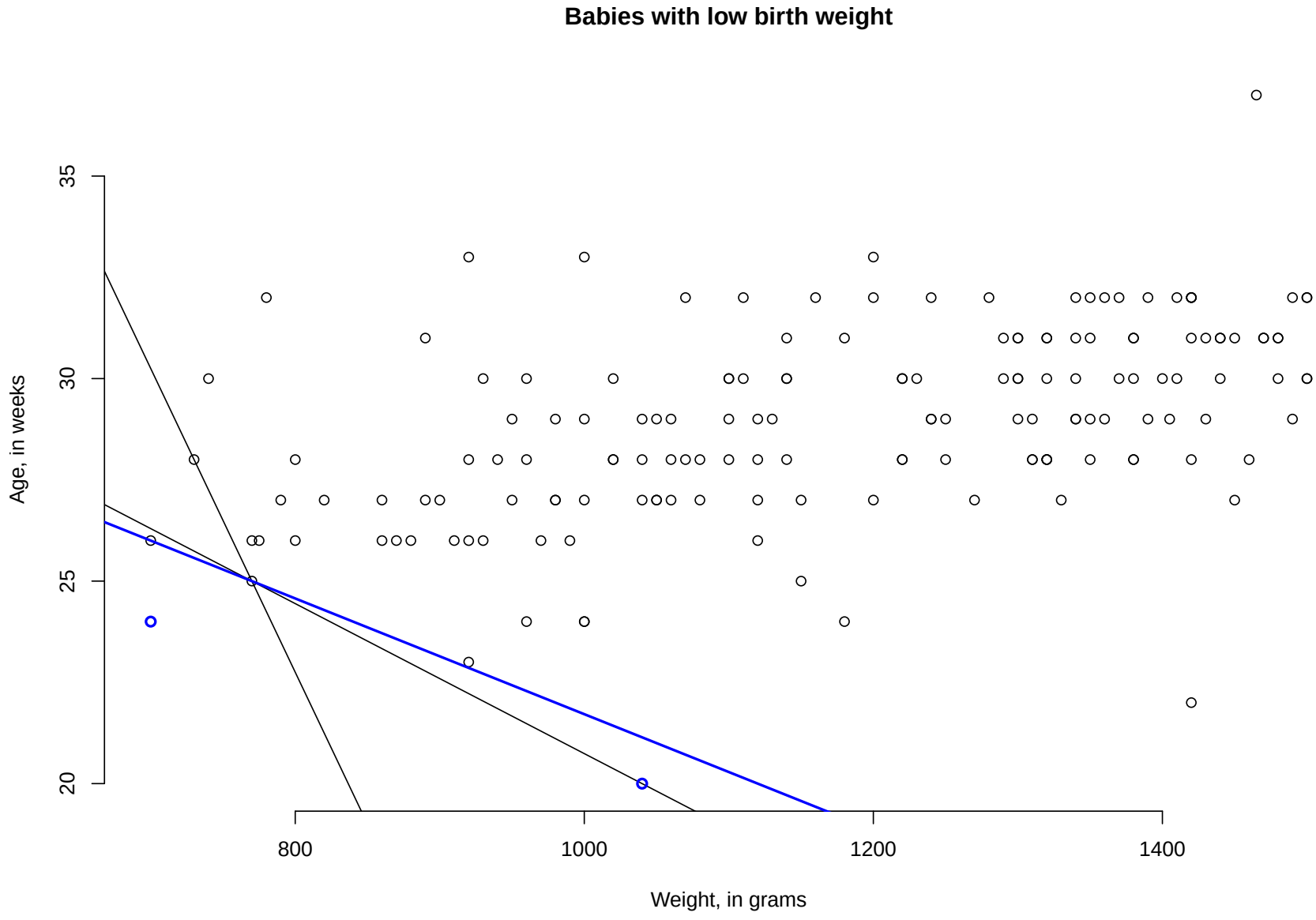
Tukey (=halfspace, location) depth-trimmed regions



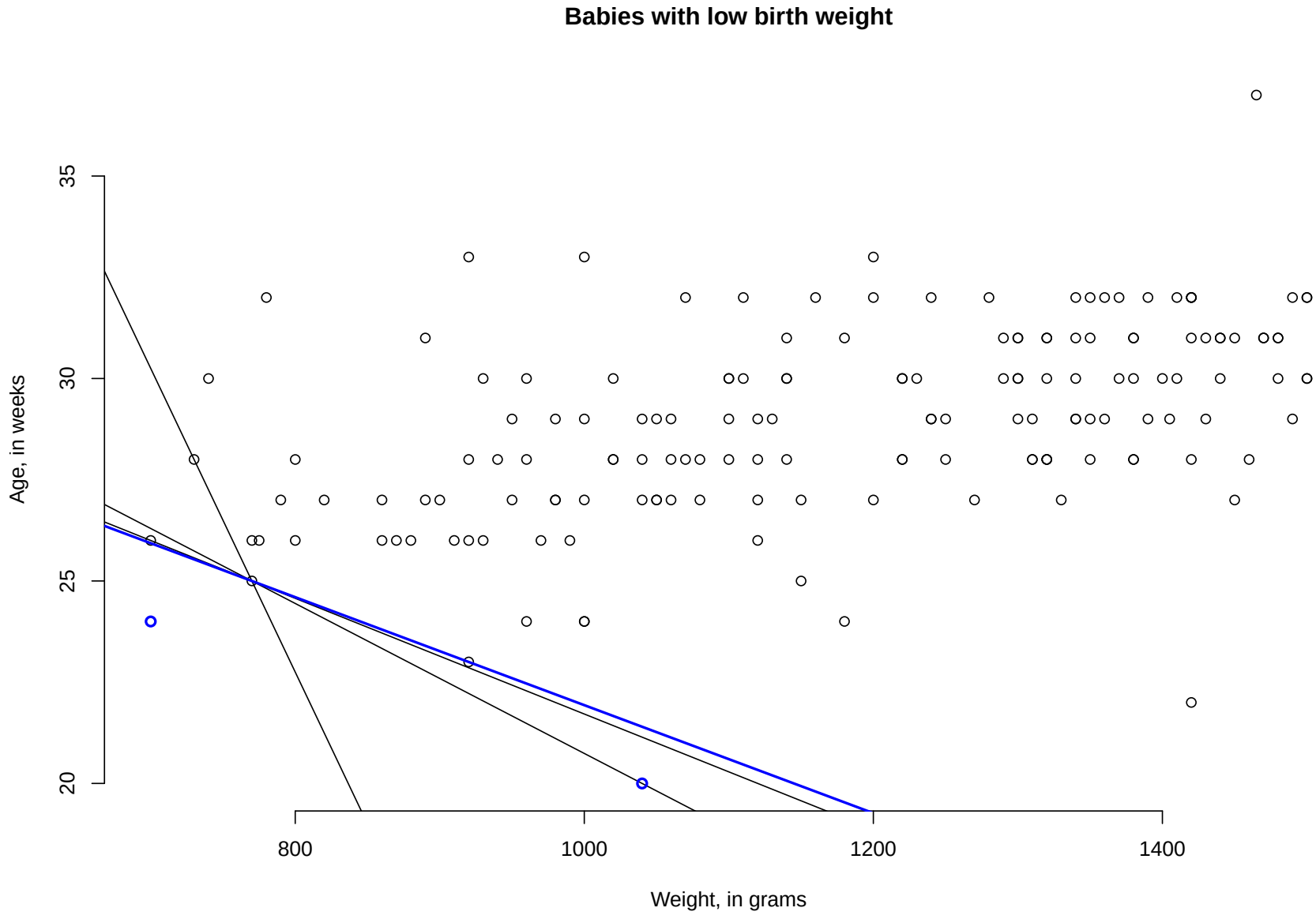
Tukey (=halfspace, location) depth-trimmed regions



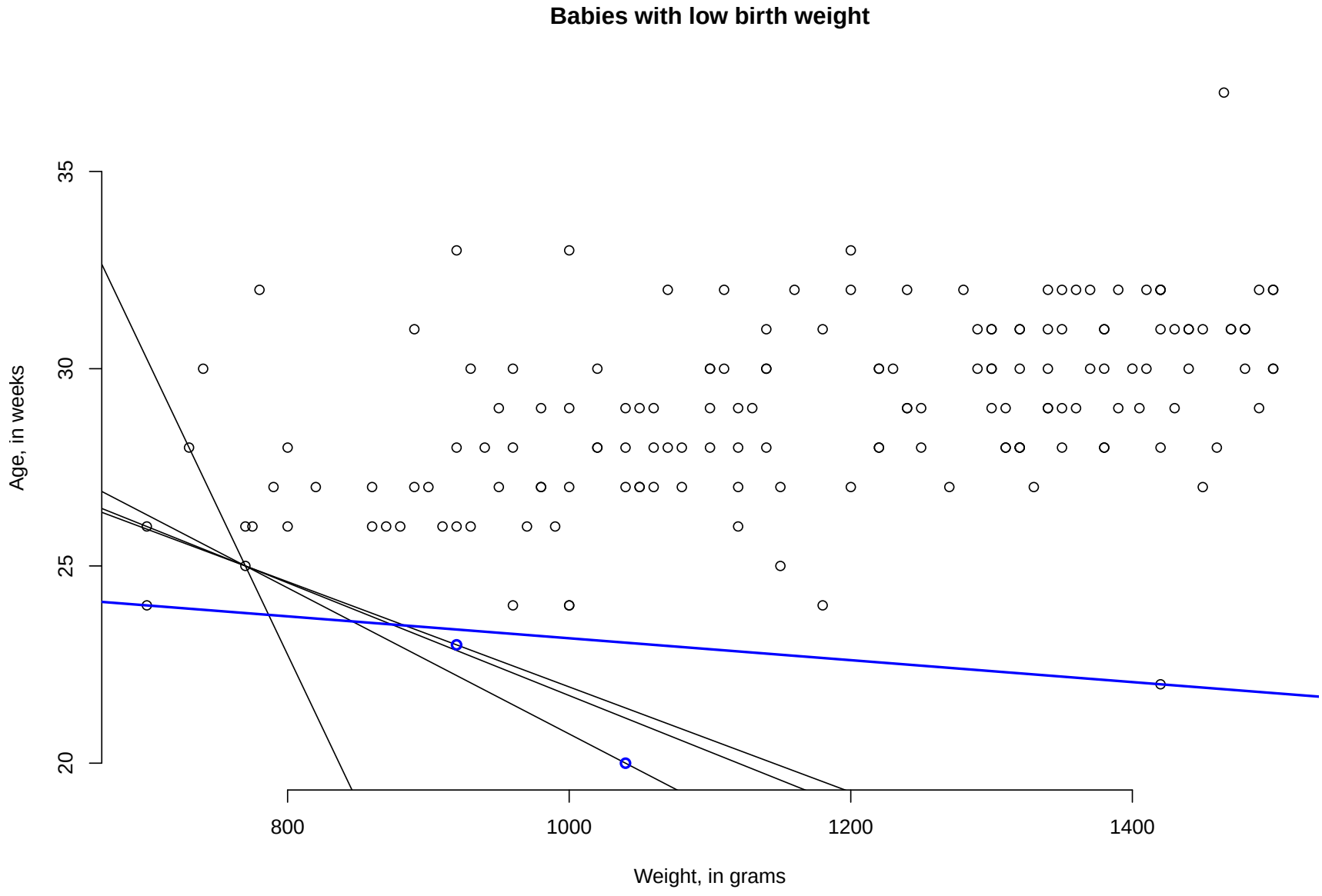
Tukey (=halfspace, location) depth-trimmed regions



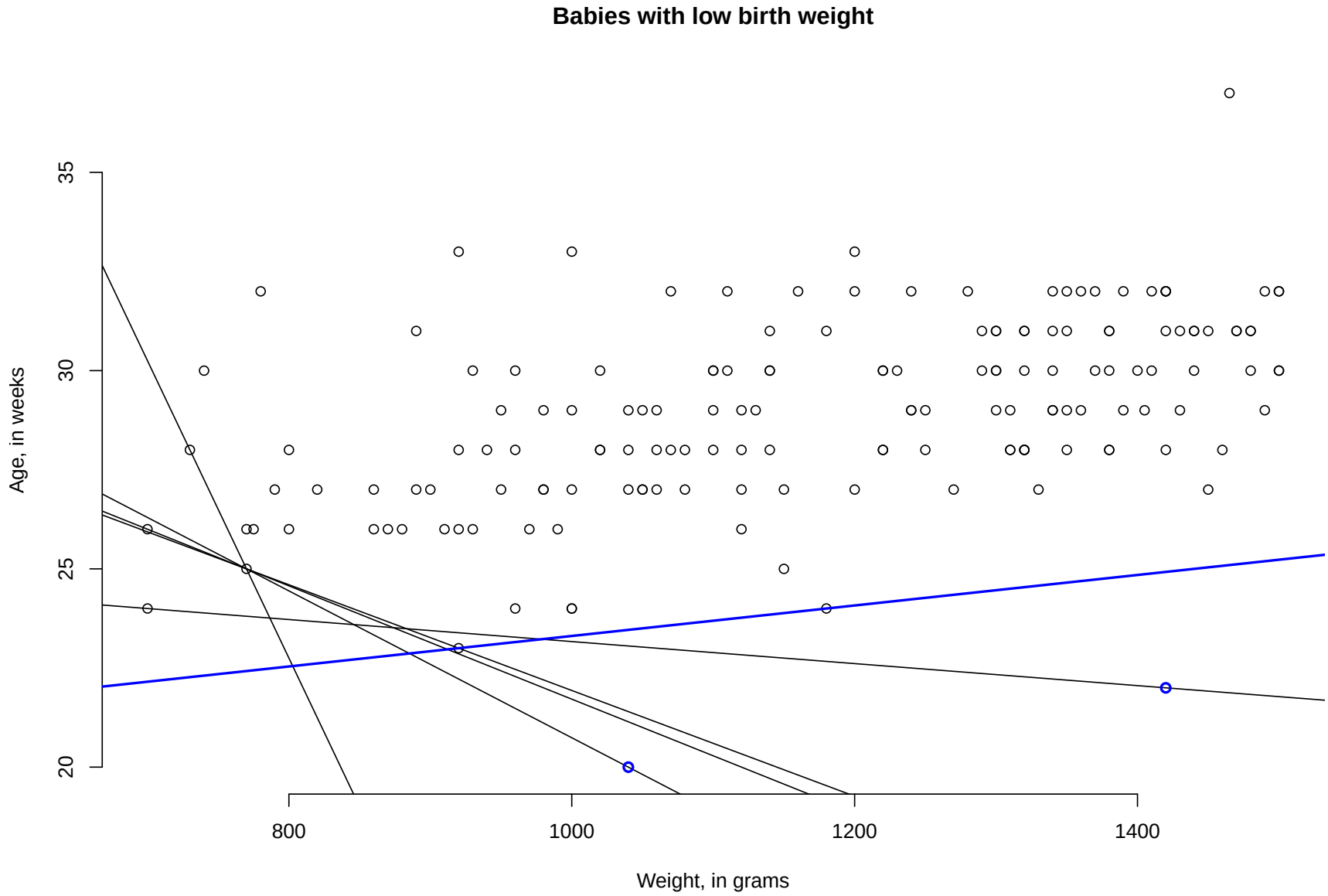
Tukey (=halfspace, location) depth-trimmed regions



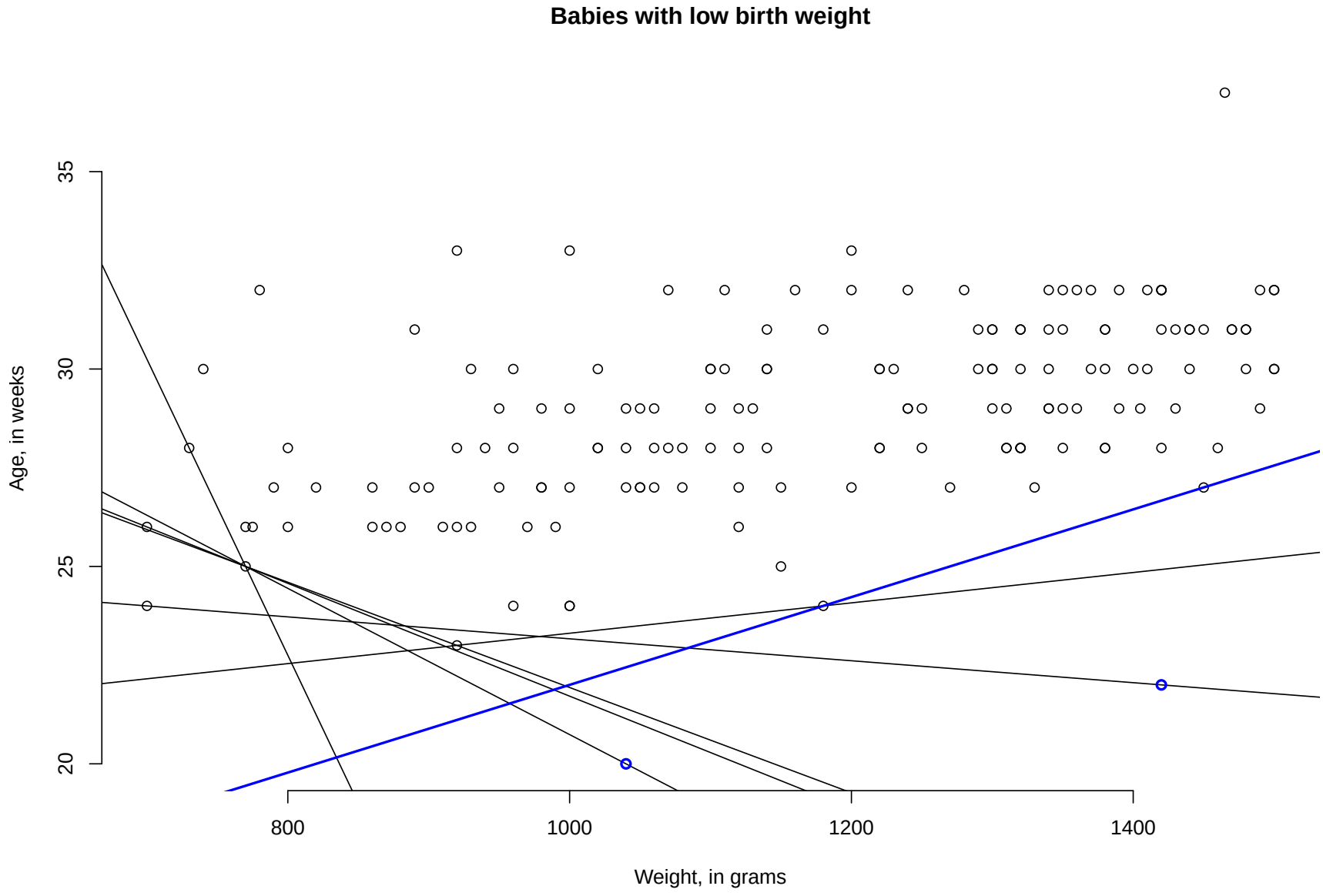
Tukey (=halfspace, location) depth-trimmed regions



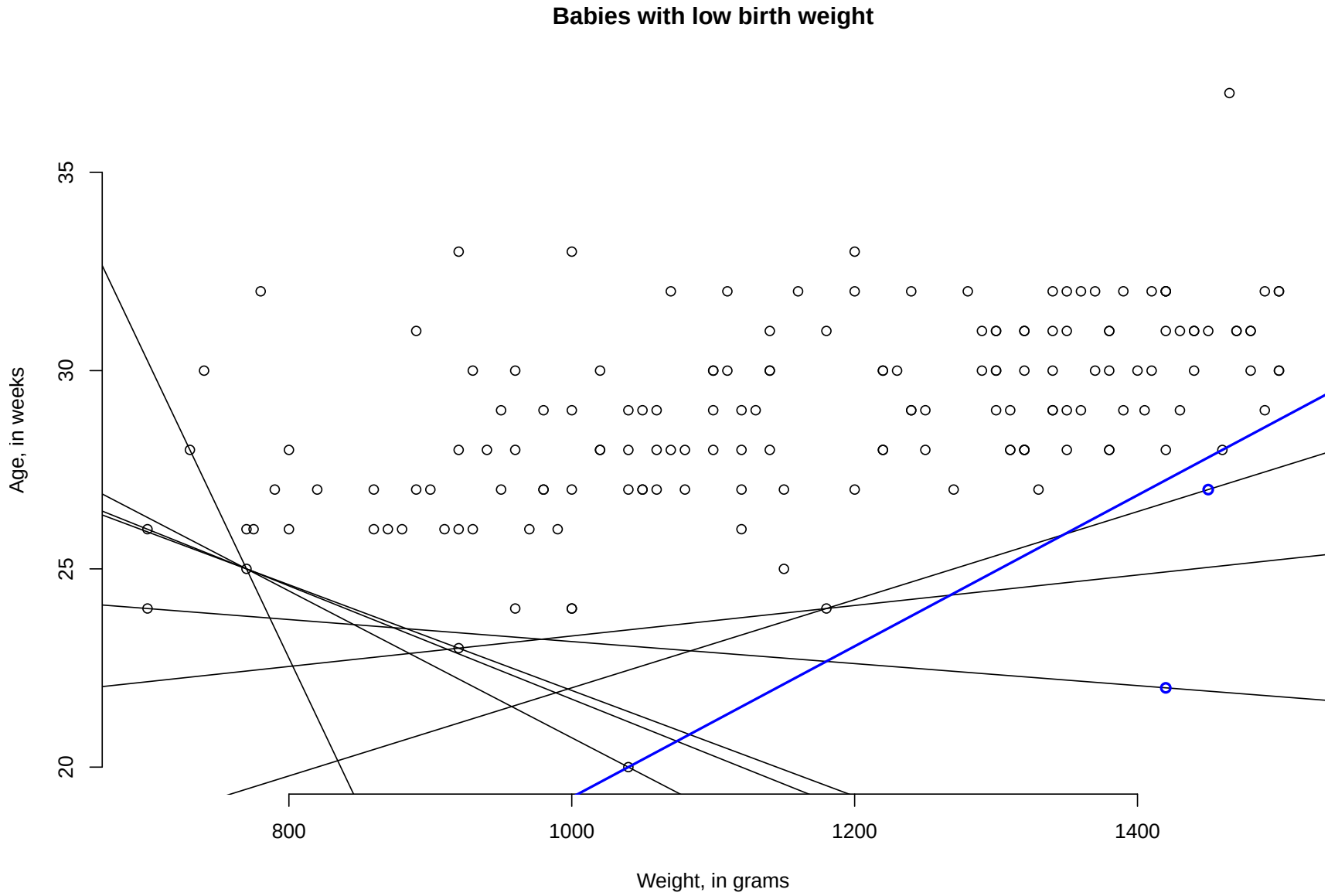
Tukey (=halfspace, location) depth-trimmed regions



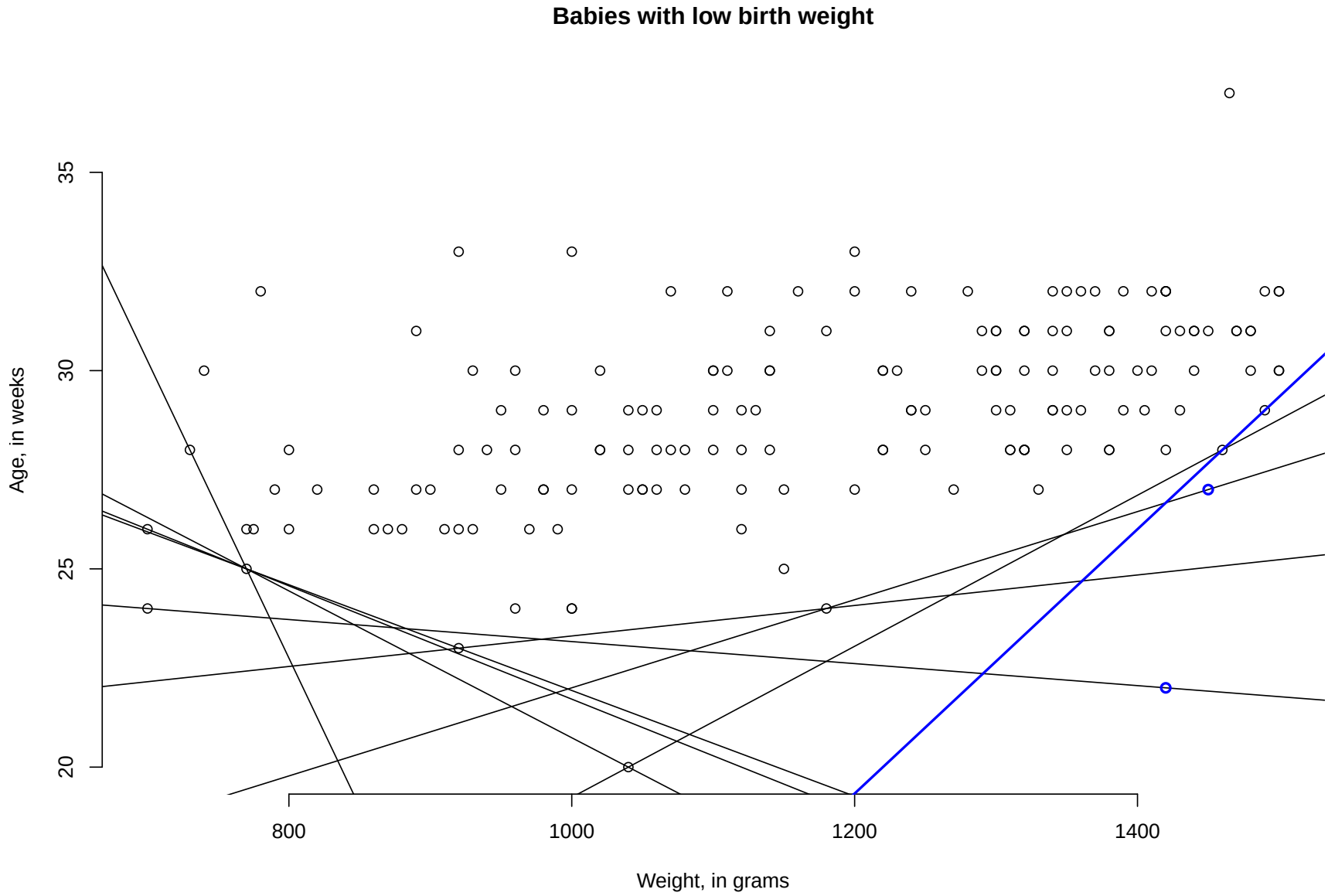
Tukey (=halfspace, location) depth-trimmed regions



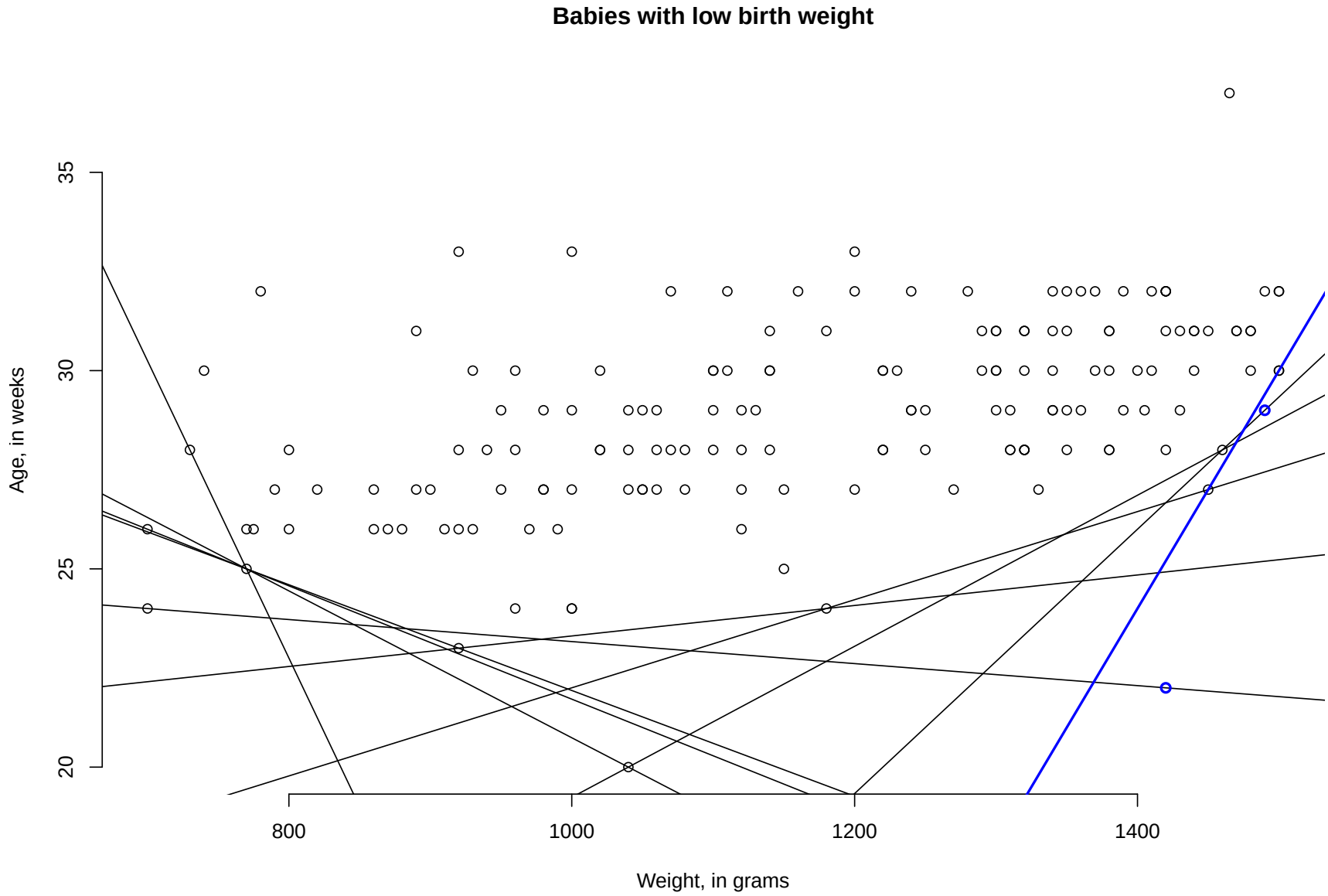
Tukey (=halfspace, location) depth-trimmed regions



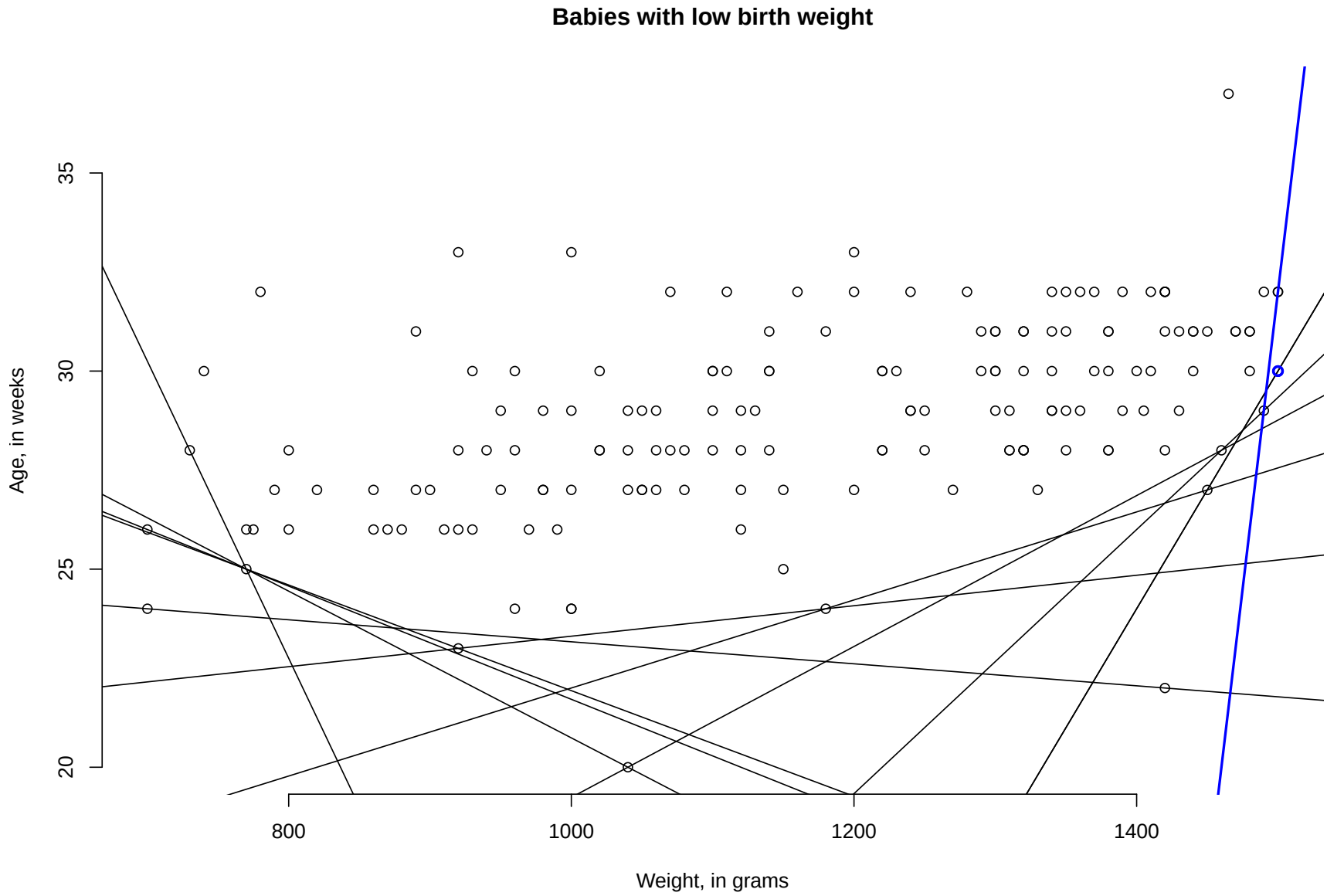
Tukey (=halfspace, location) depth-trimmed regions



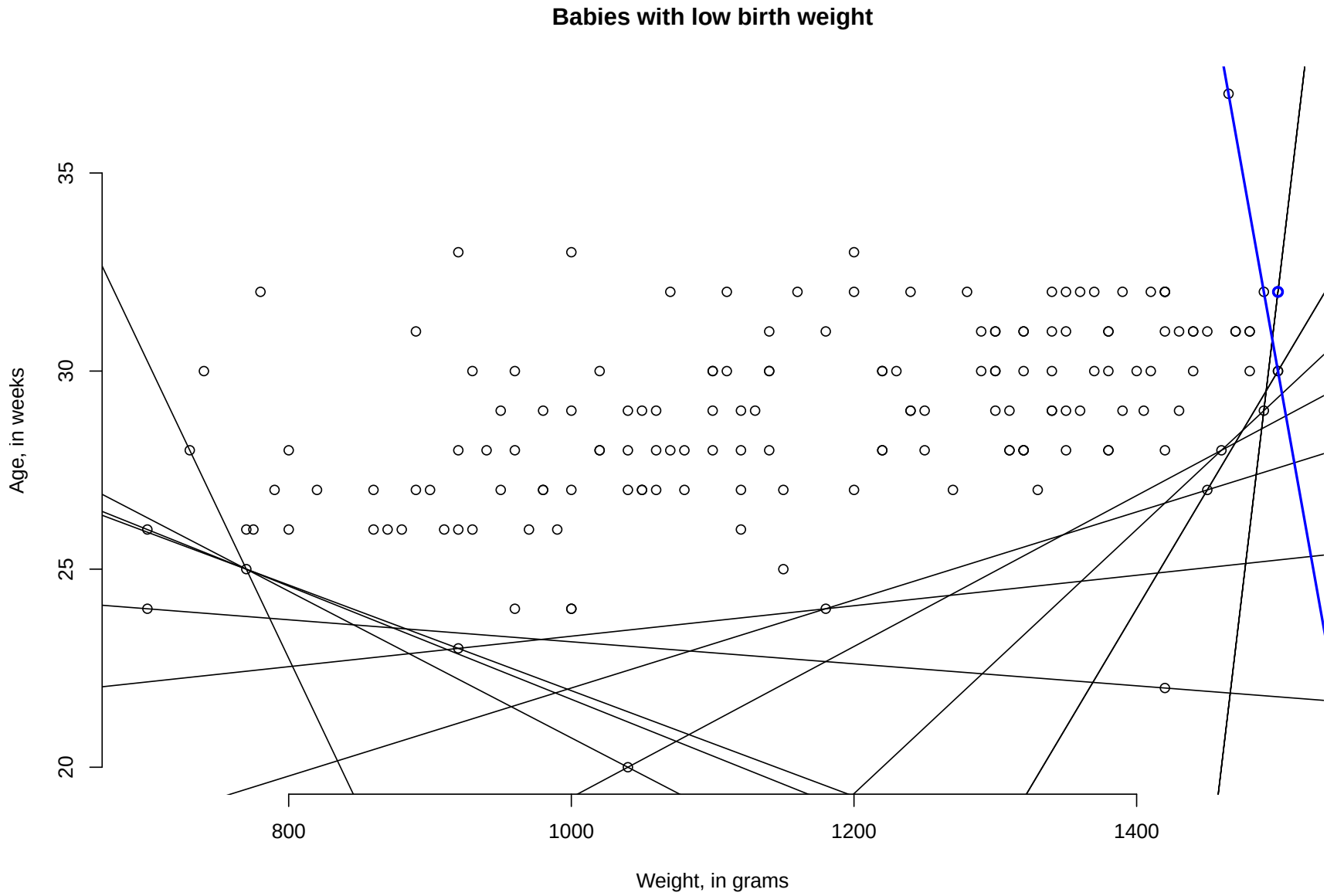
Tukey (=halfspace, location) depth-trimmed regions



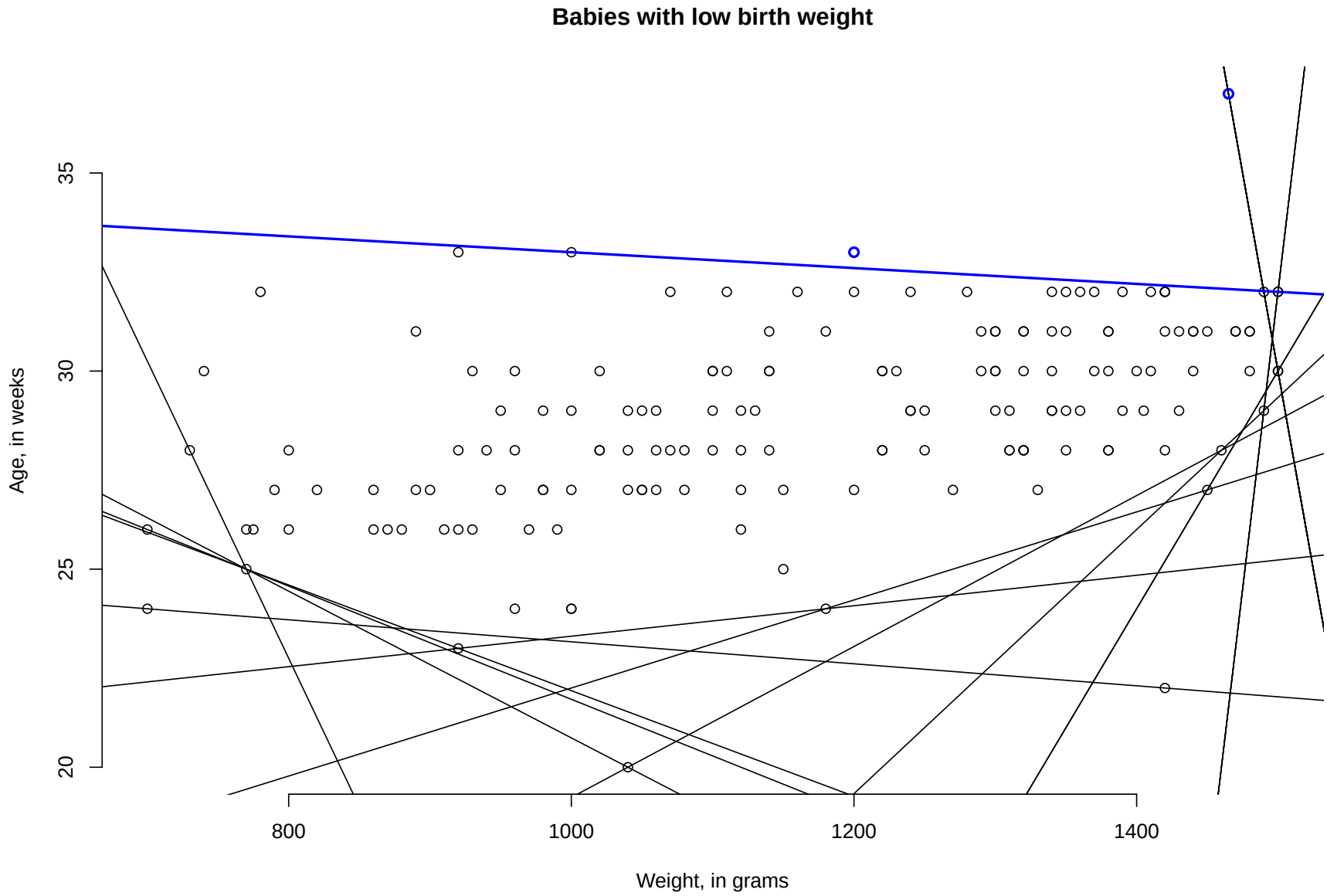
Tukey (=halfspace, location) depth-trimmed regions



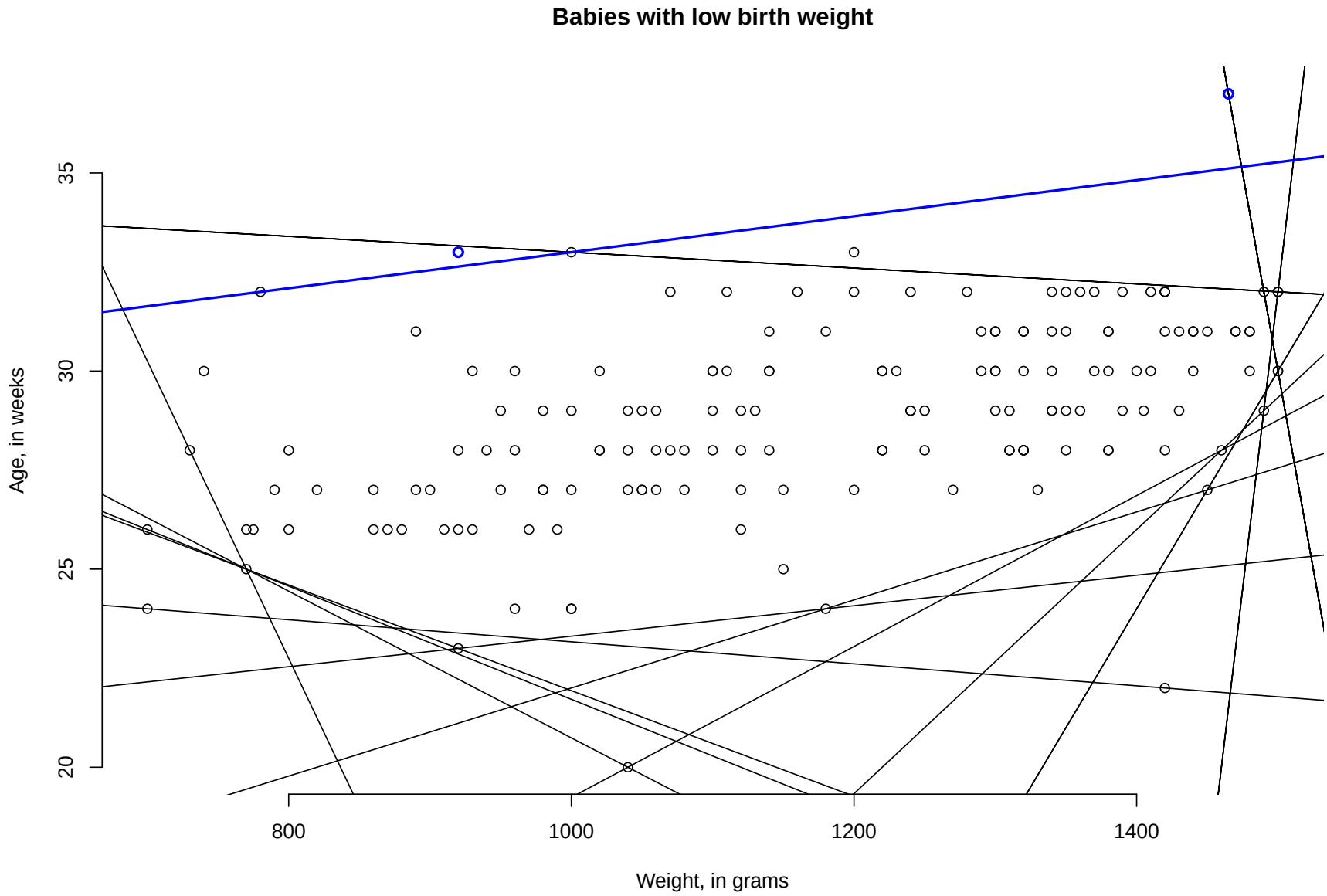
Tukey (=halfspace, location) depth-trimmed regions



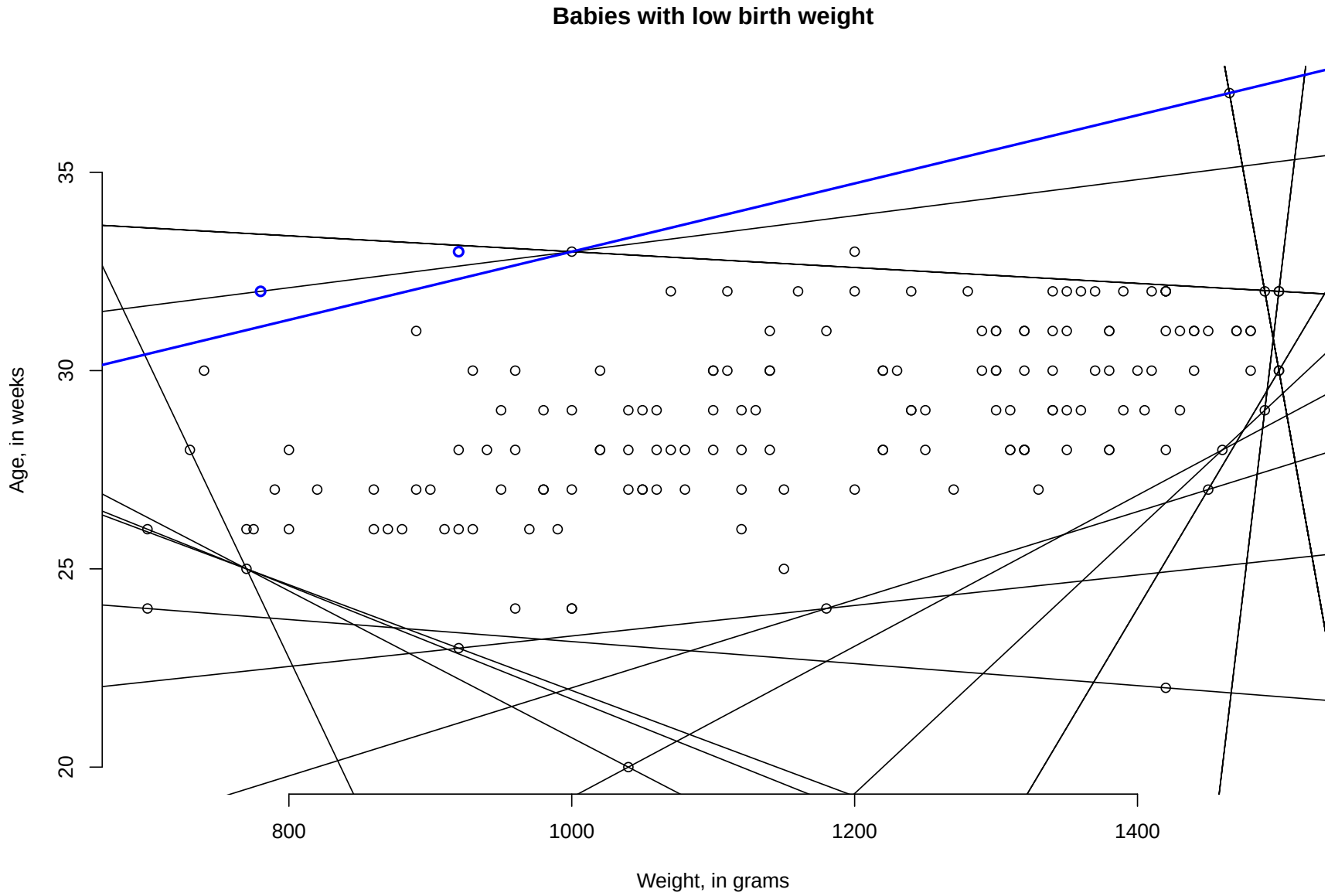
Tukey (=halfspace, location) depth-trimmed regions



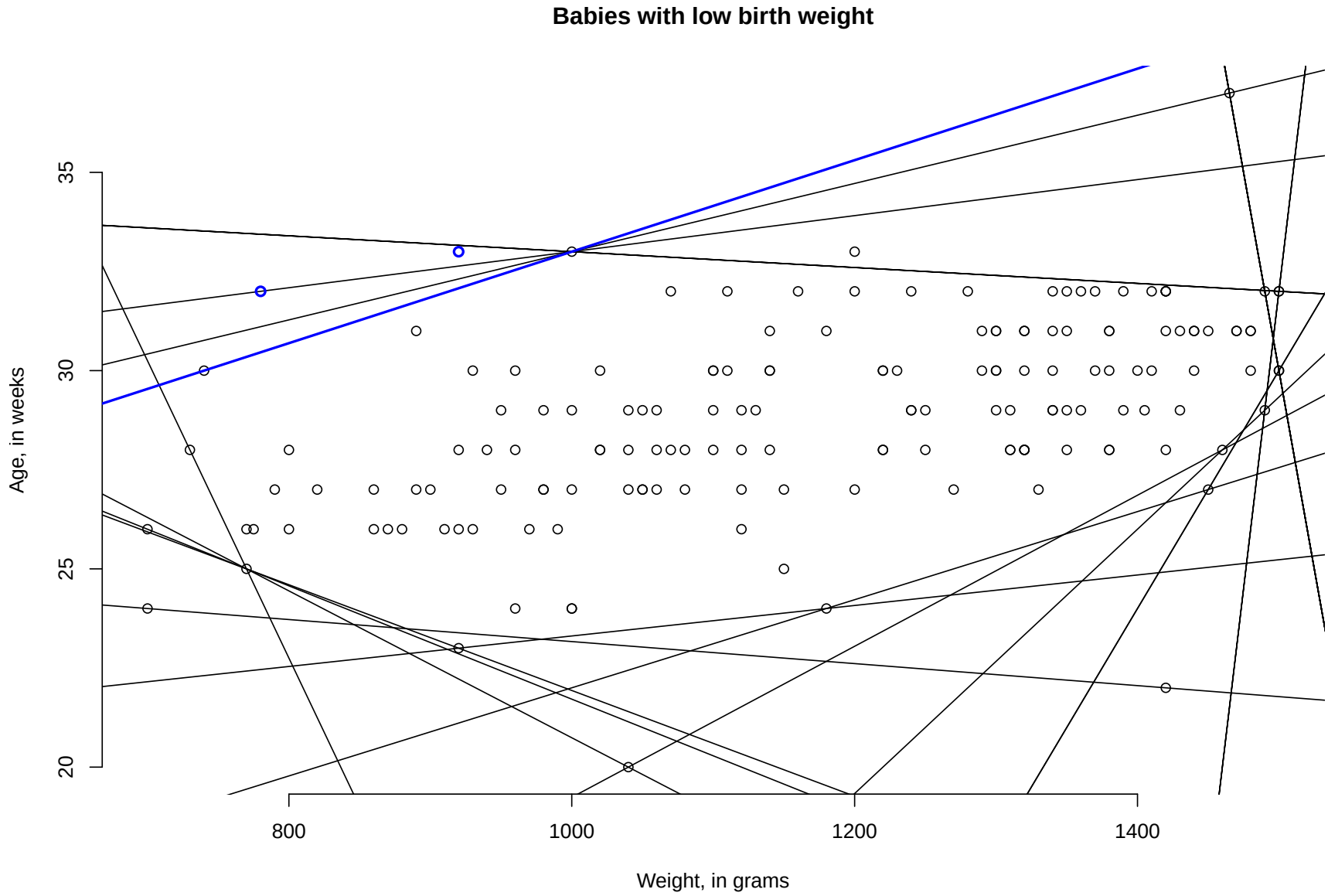
Tukey (=halfspace, location) depth-trimmed regions



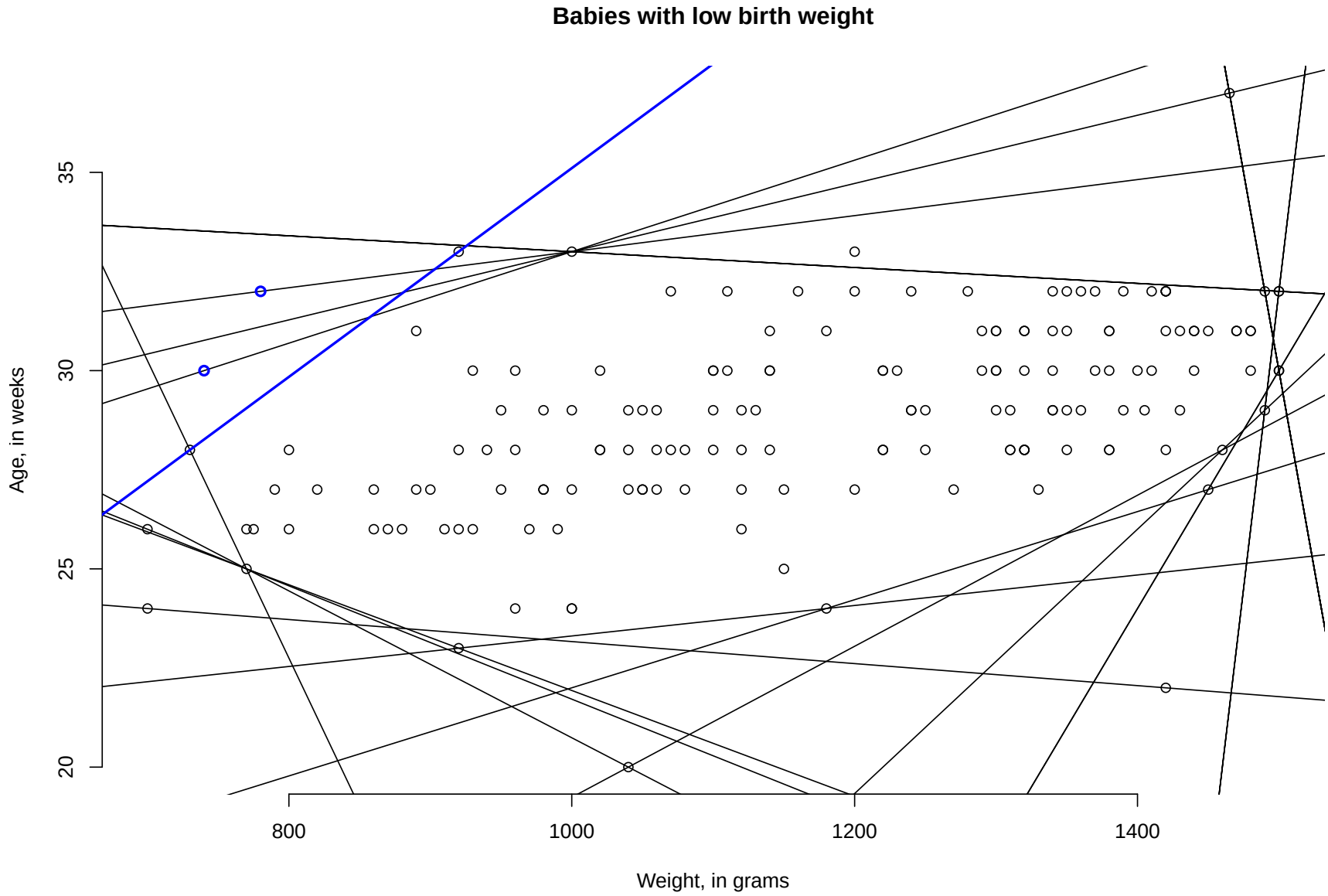
Tukey (=halfspace, location) depth-trimmed regions



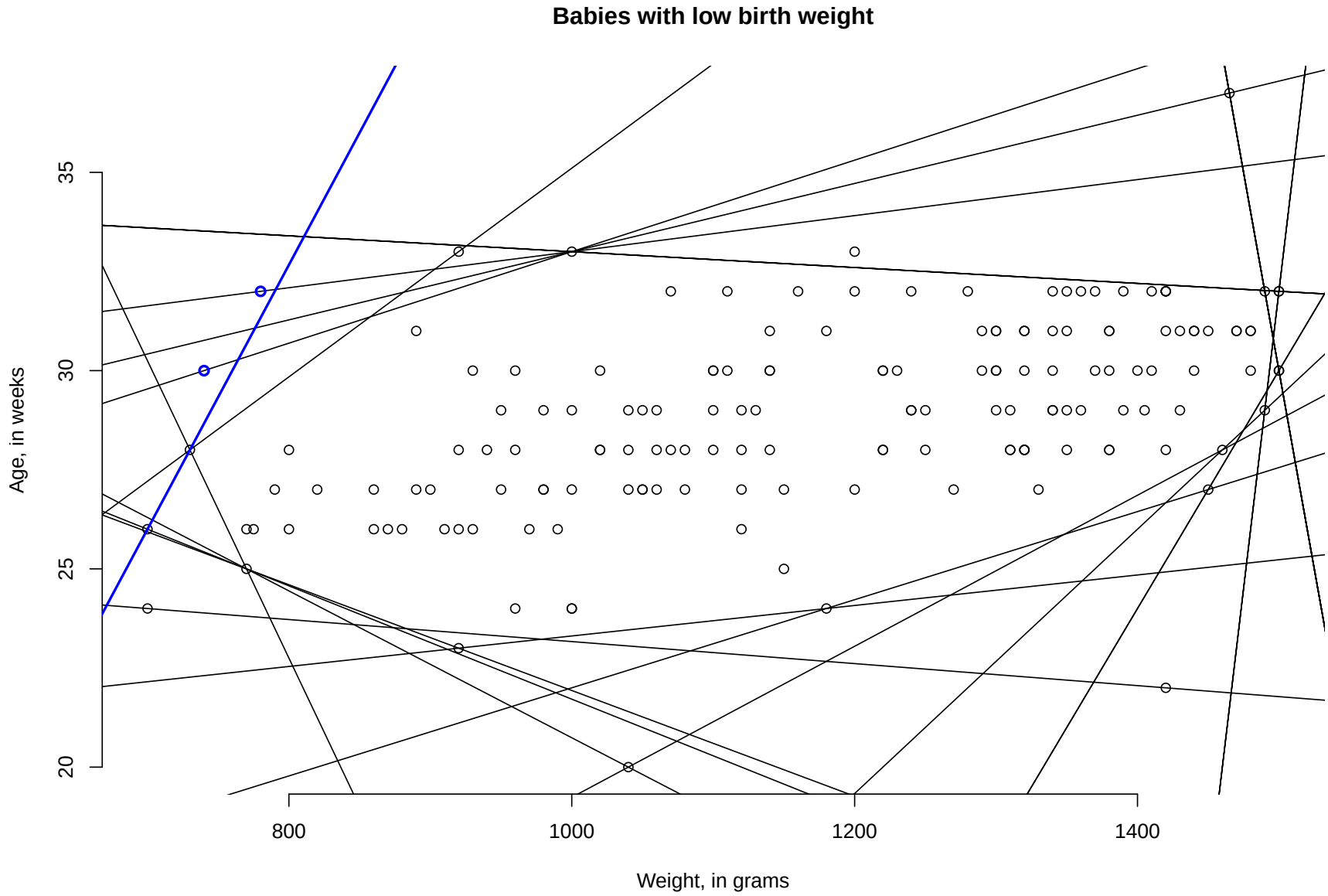
Tukey (=halfspace, location) depth-trimmed regions



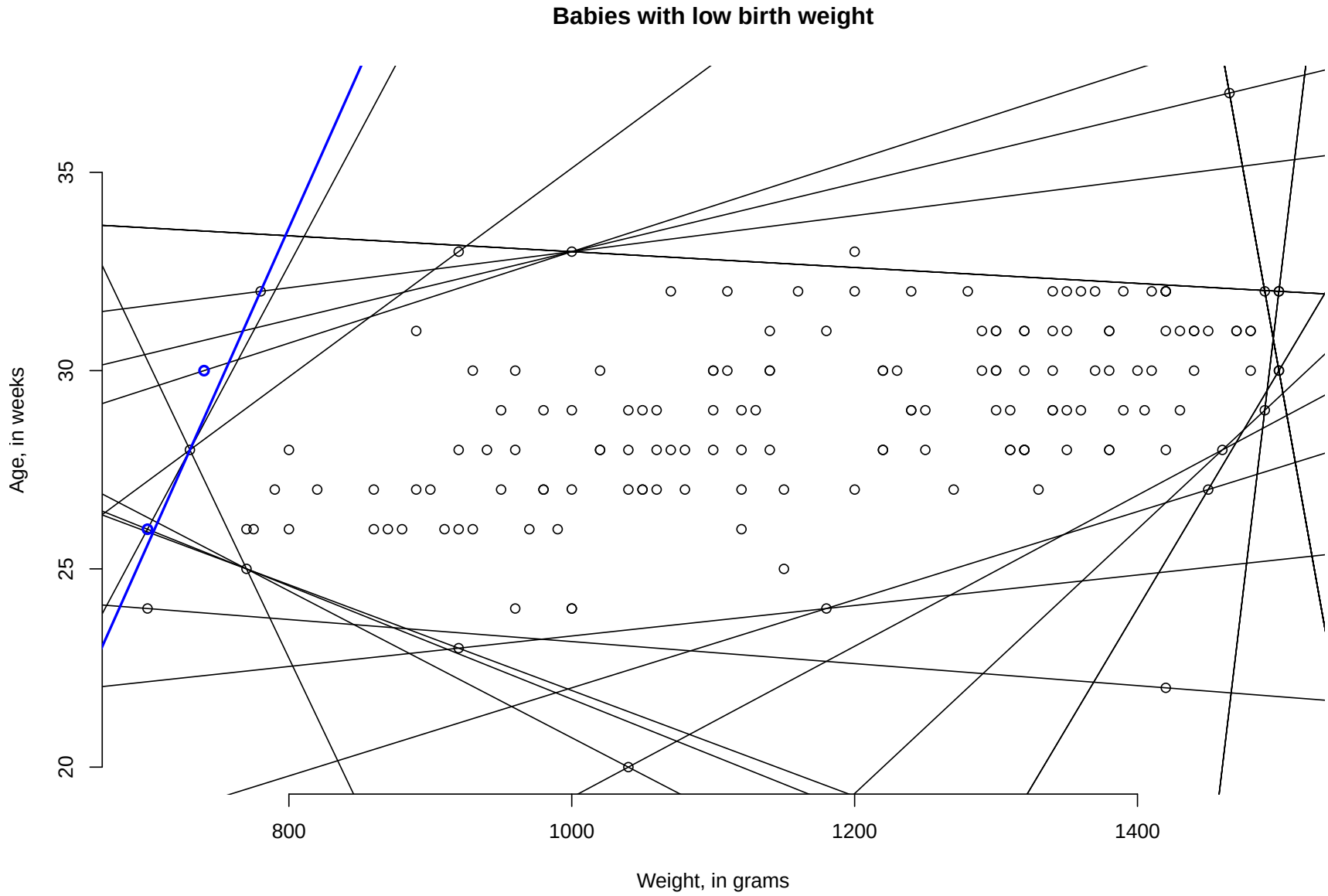
Tukey (=halfspace, location) depth-trimmed regions



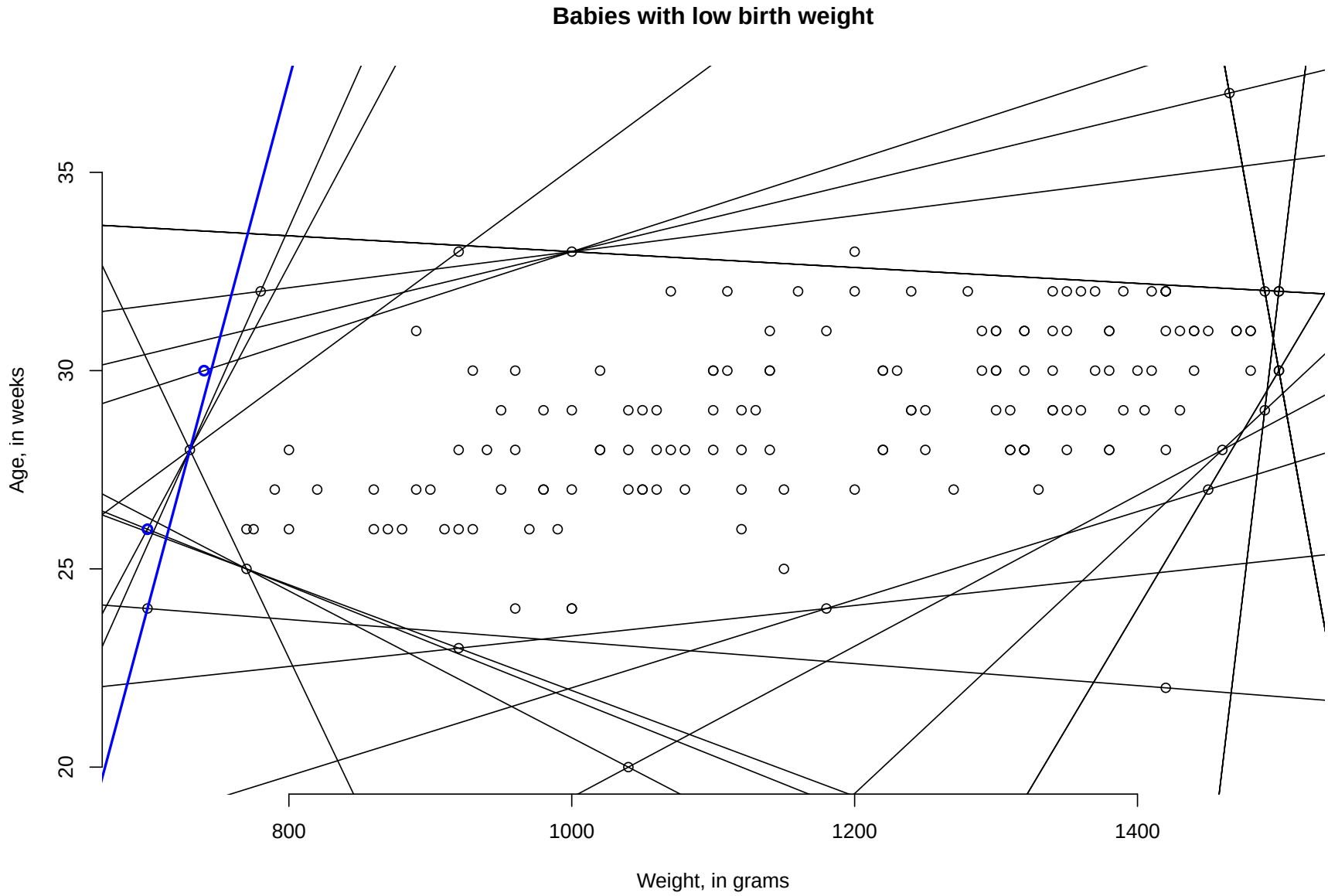
Tukey (=halfspace, location) depth-trimmed regions



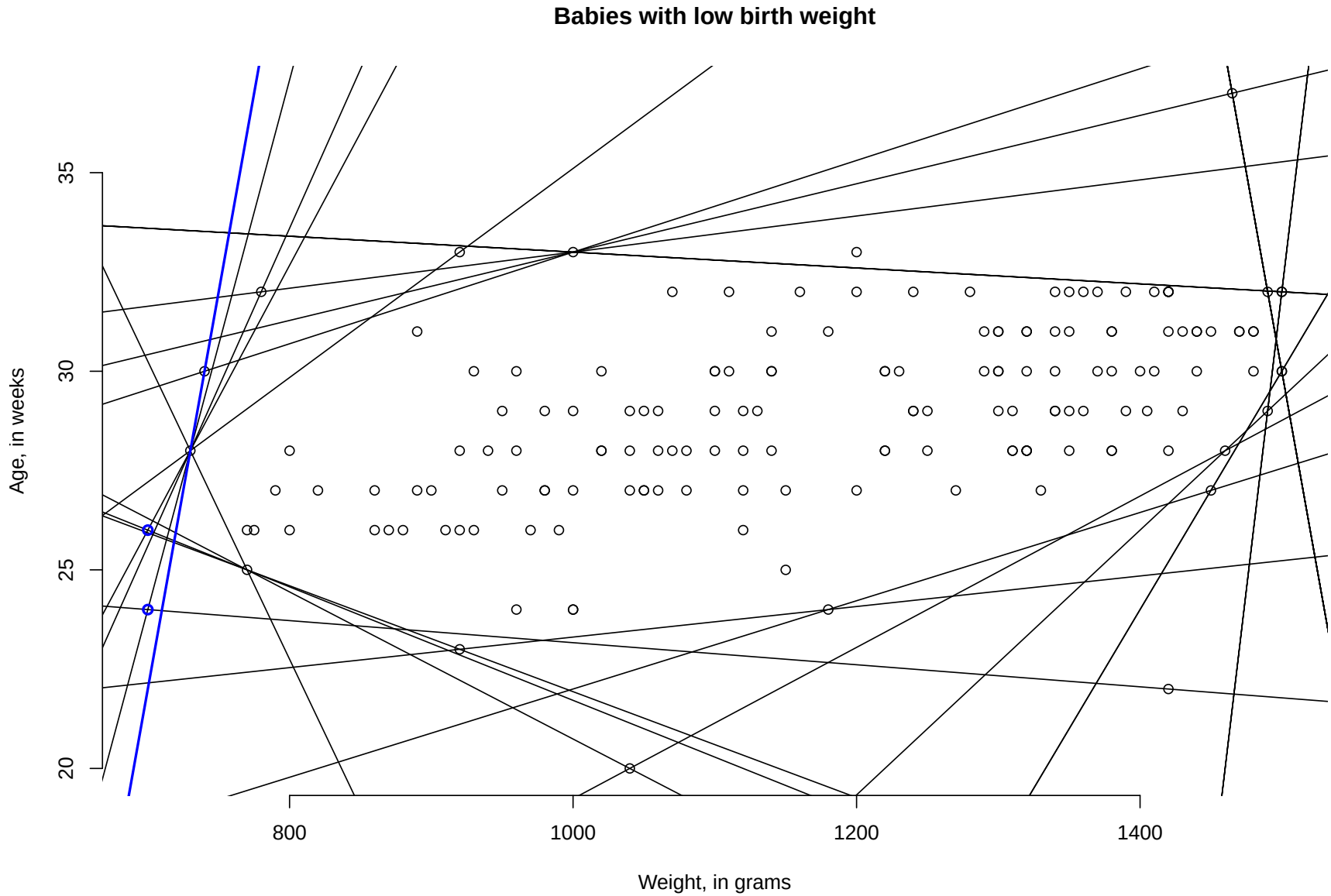
Tukey (=halfspace, location) depth-trimmed regions



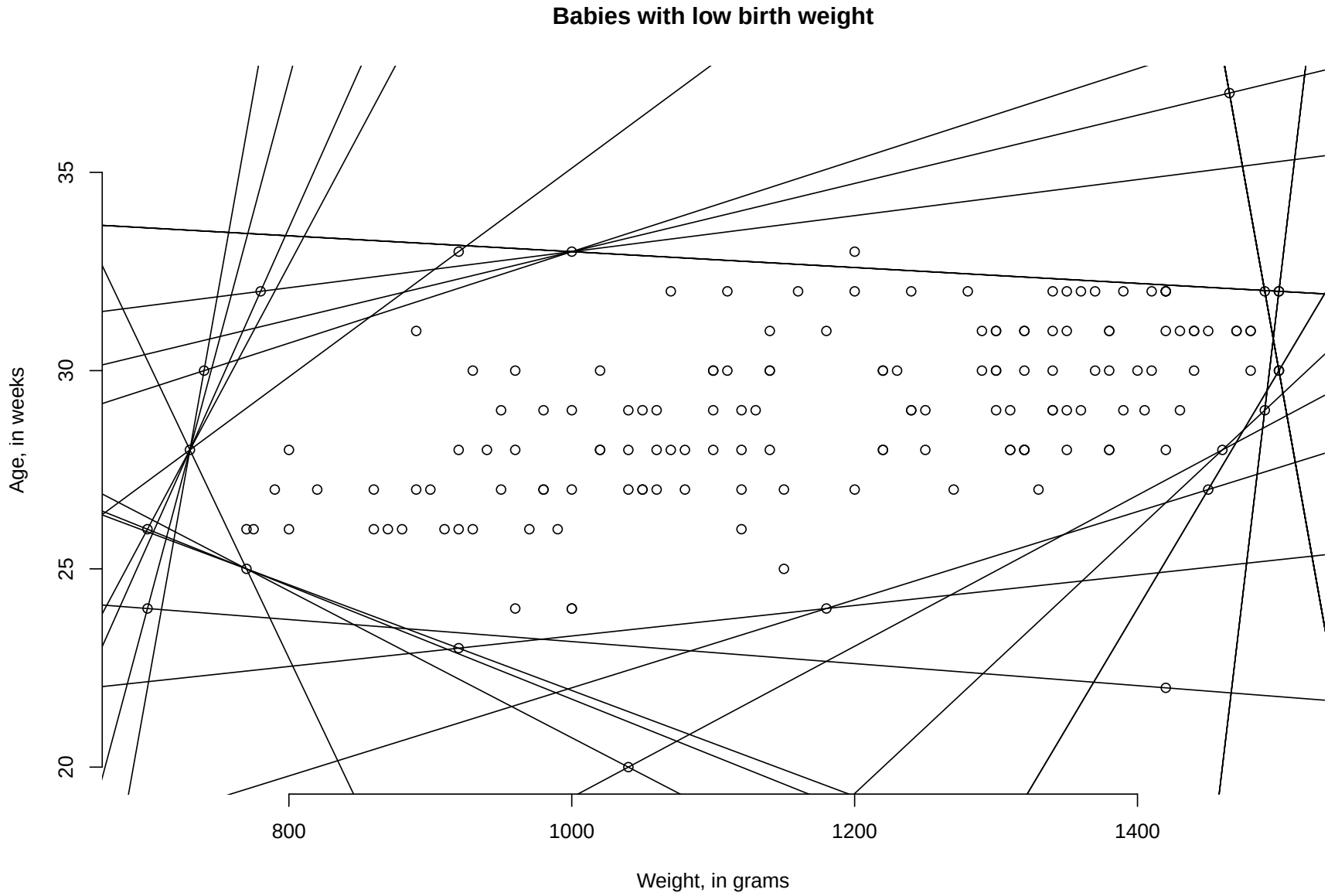
Tukey (=halfspace, location) depth-trimmed regions



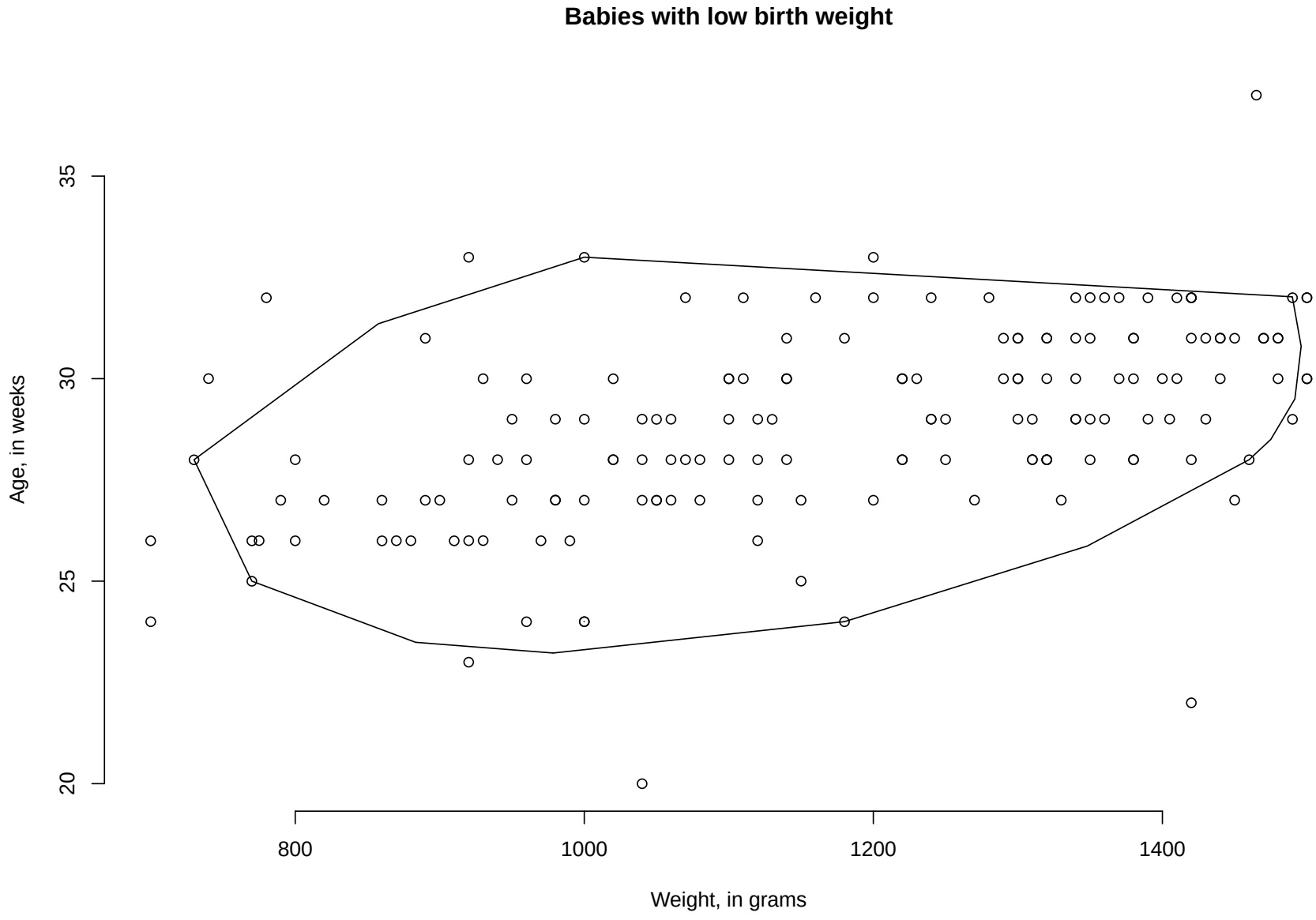
Tukey (=halfspace, location) depth-trimmed regions



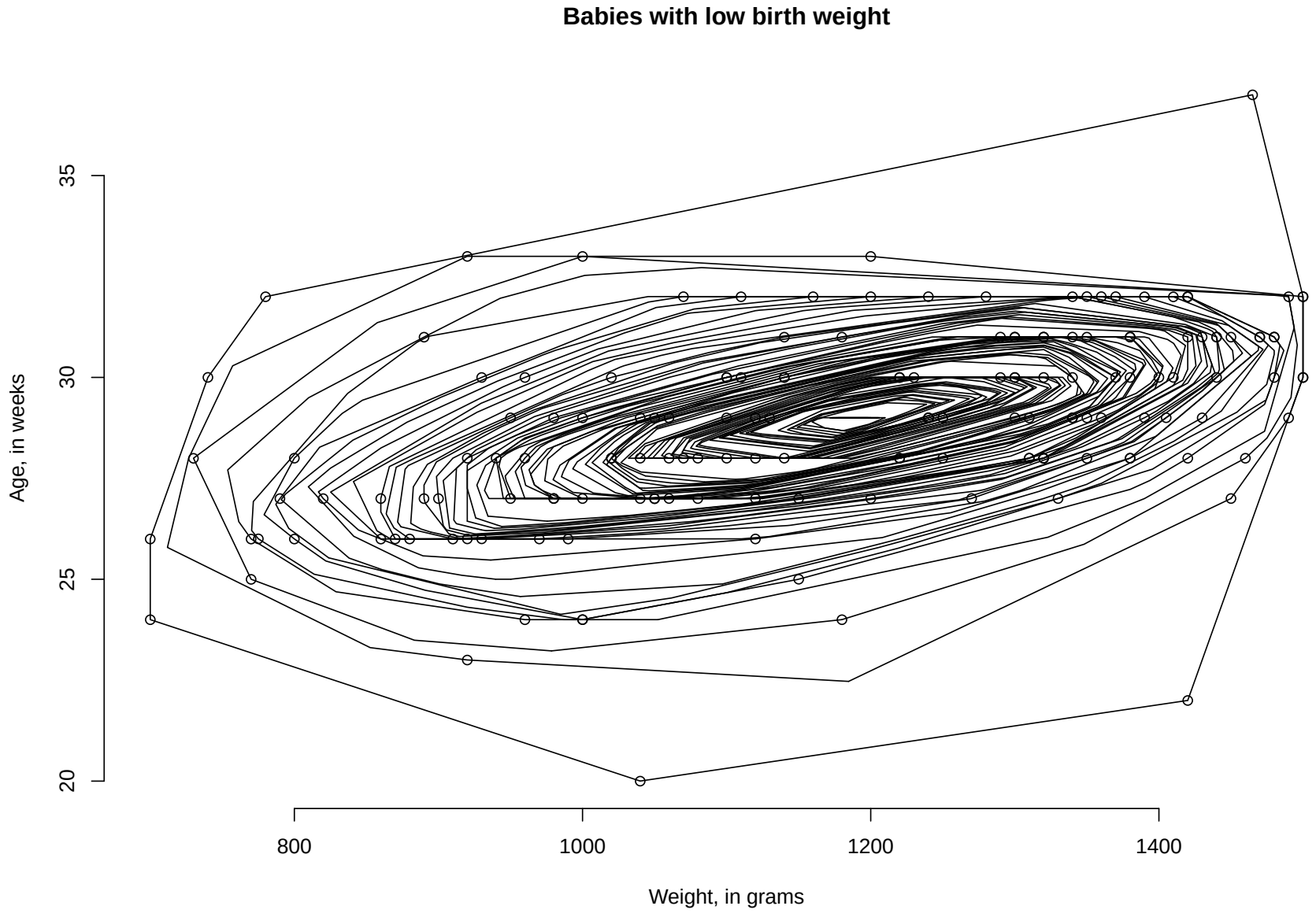
Tukey (=halfspace, location) depth-trimmed regions



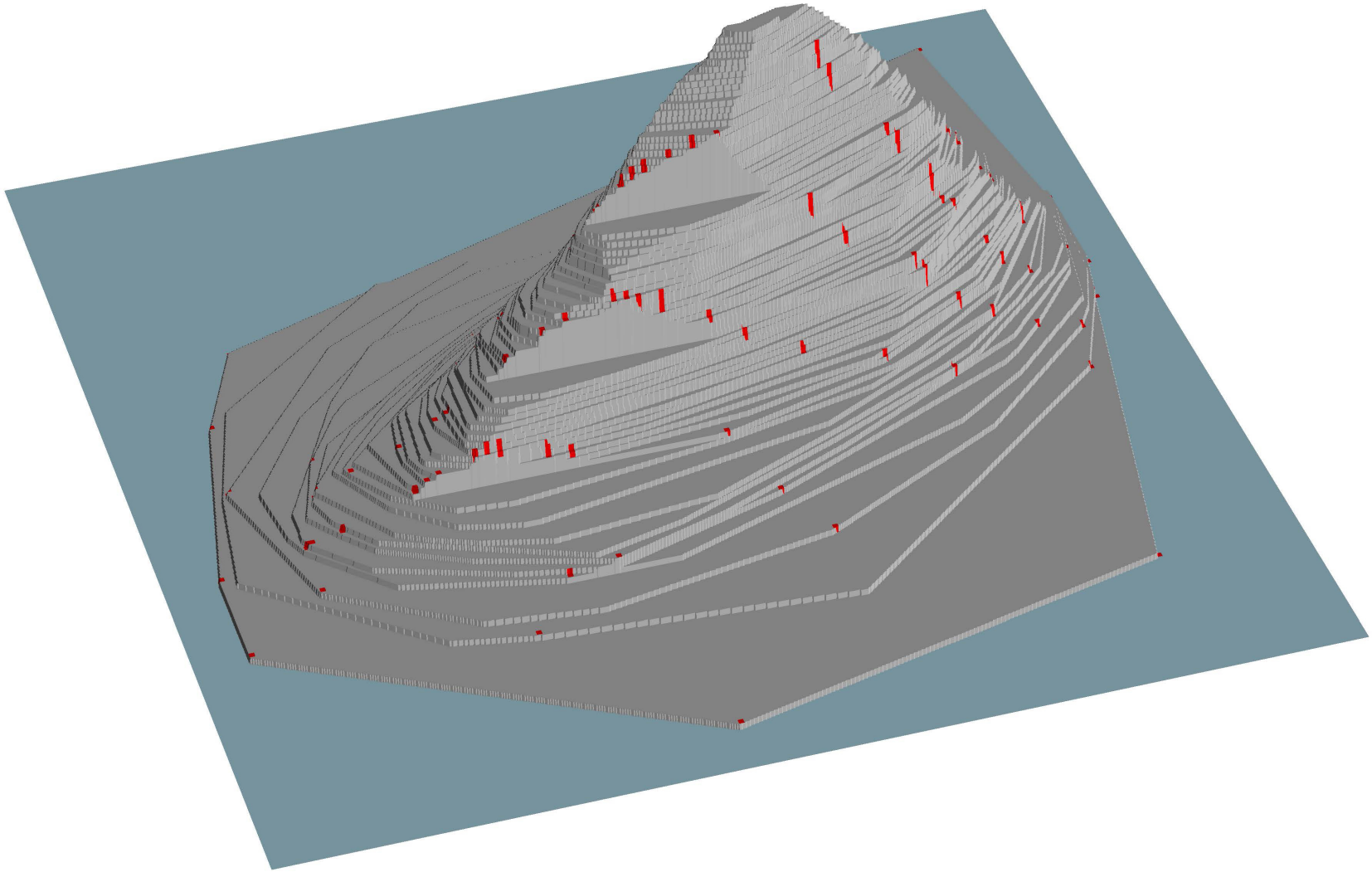
Tukey (=halfspace, location) depth-trimmed regions



Tukey (=halfspace, location) depth-trimmed regions



Tukey (=halfspace, location) data depth



Directional quantiles

Kong, Mizera (2008)

For a random vector $X \in \mathbb{R}^d$ define the $\boldsymbol{u}$-directional α-th quantile:

$$Q(\alpha, \boldsymbol{u}, X) = \inf\{x \,:\, P(\boldsymbol{u}^T X \leq x) \geq \alpha\}.$$

Directional quantile envelope:

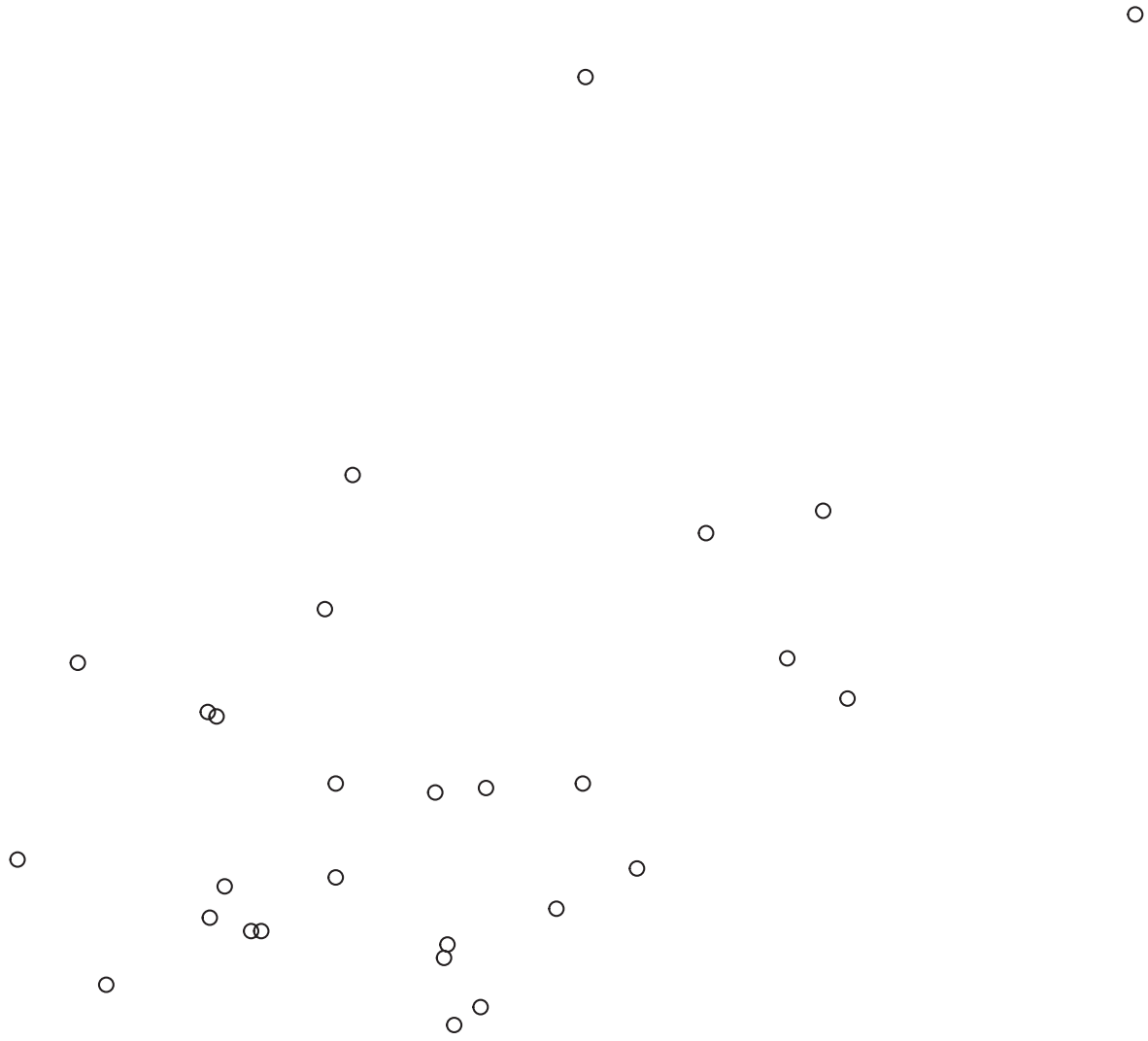
$$R_\alpha(X) = \bigcap_{\boldsymbol{u} \in S^{d-1}} H\big(\boldsymbol{u}, Q(\alpha, \boldsymbol{u}, X)\big),$$

where $H(\boldsymbol{u}, q) = \{x : \boldsymbol{u}^T x \geq q\}$ is the *supporting halfspace* determined by $\boldsymbol{u} \in S^{d-1}$ and $q \in \mathbb{R}$.

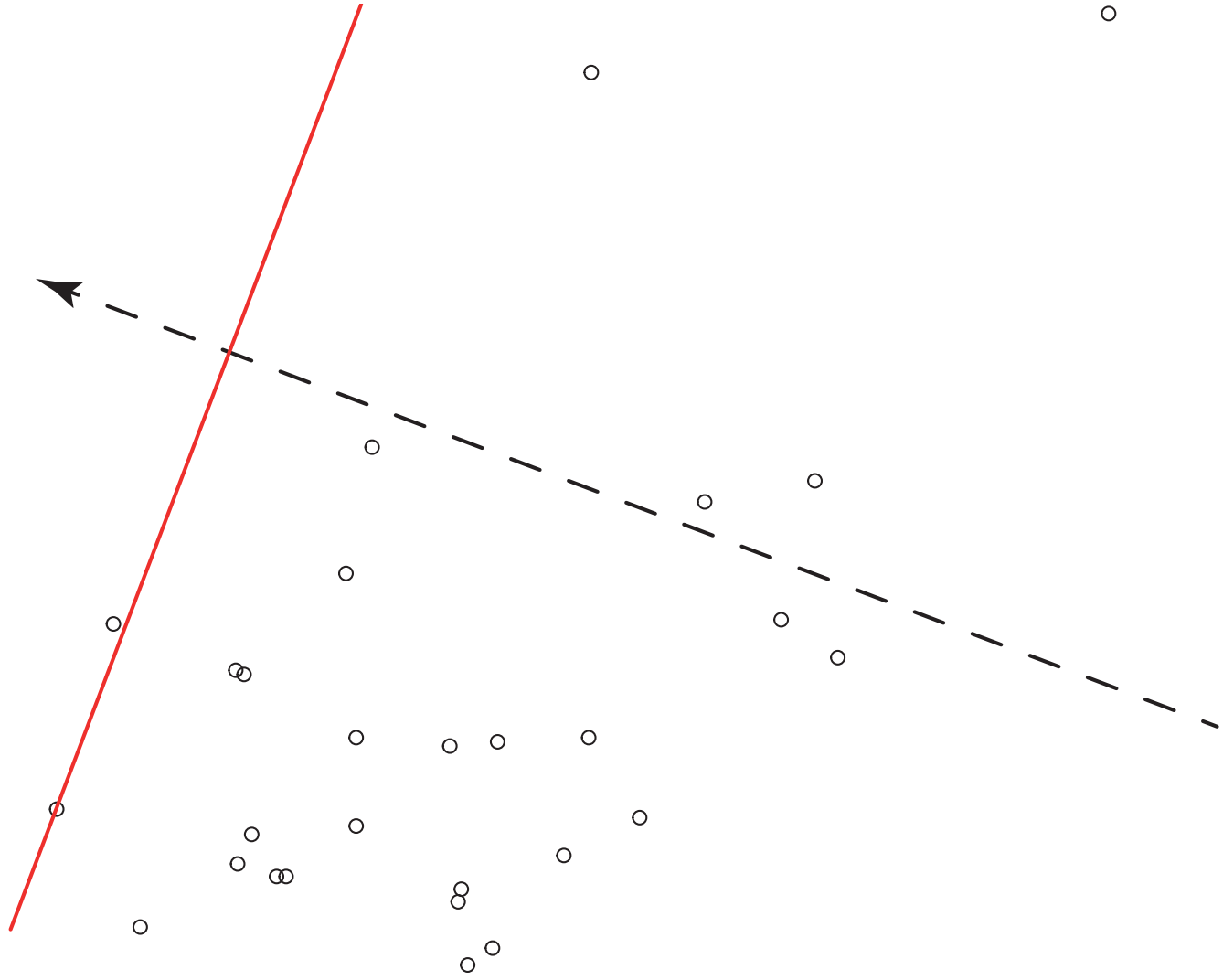
Then for every $p \in (0; \frac{1}{2}]$ directional quantile envelopes coincide with the Tukey regions:

$$R_\alpha(X) = D_\alpha^{Tuk}(X) \text{ for every } \alpha \in (0; \frac{1}{2}].$$

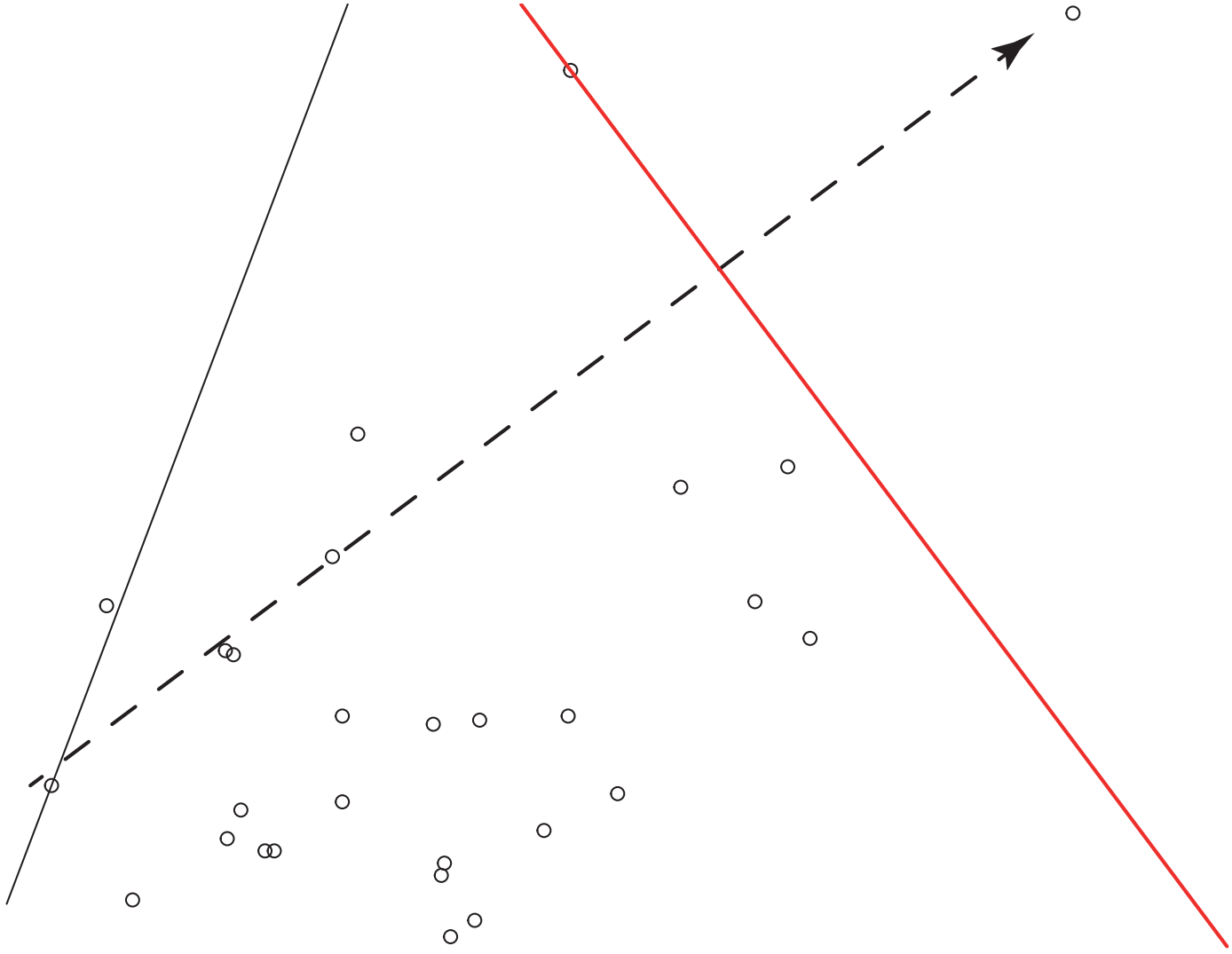
Directional quantiles (illustration)



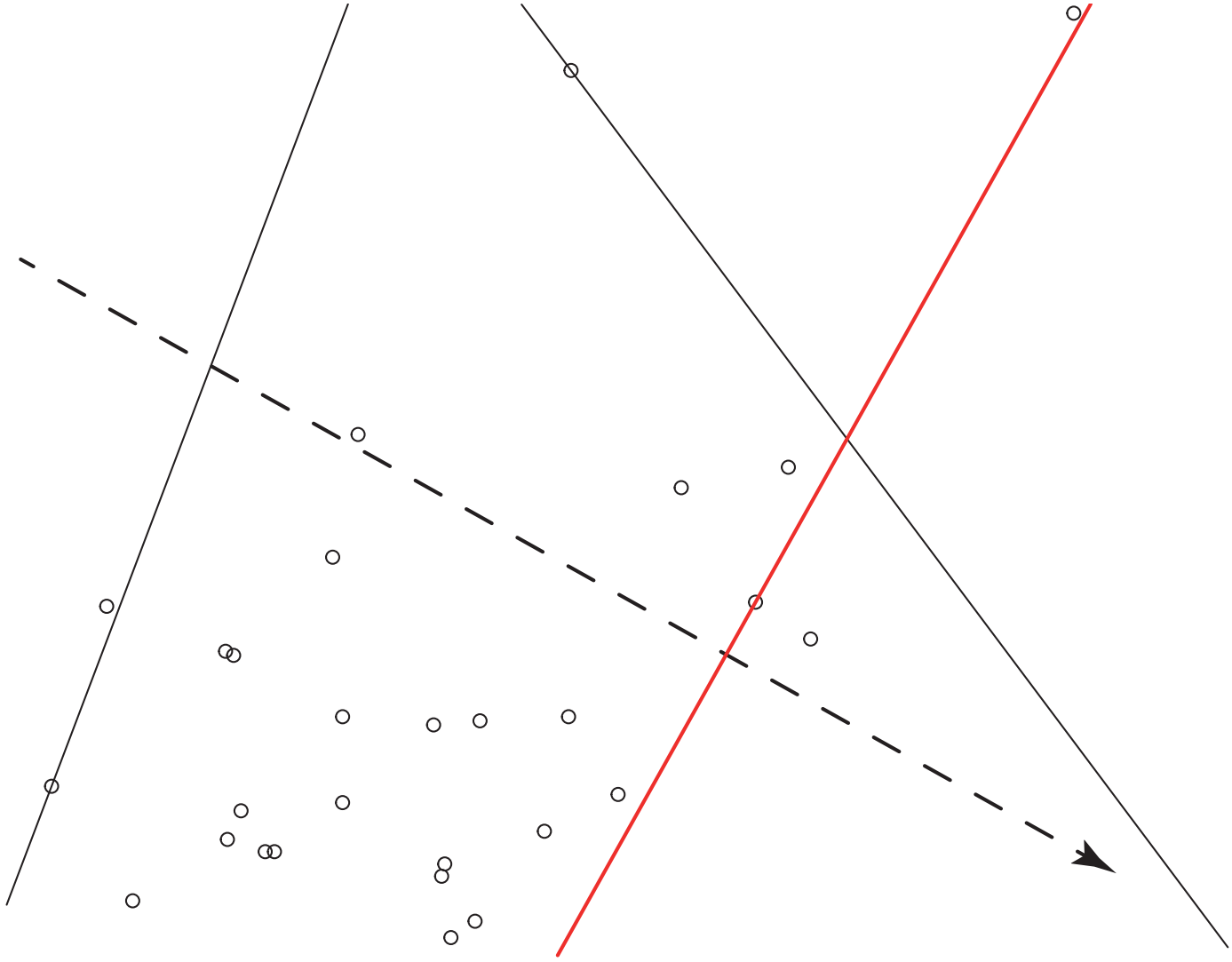
Directional quantiles (illustration)



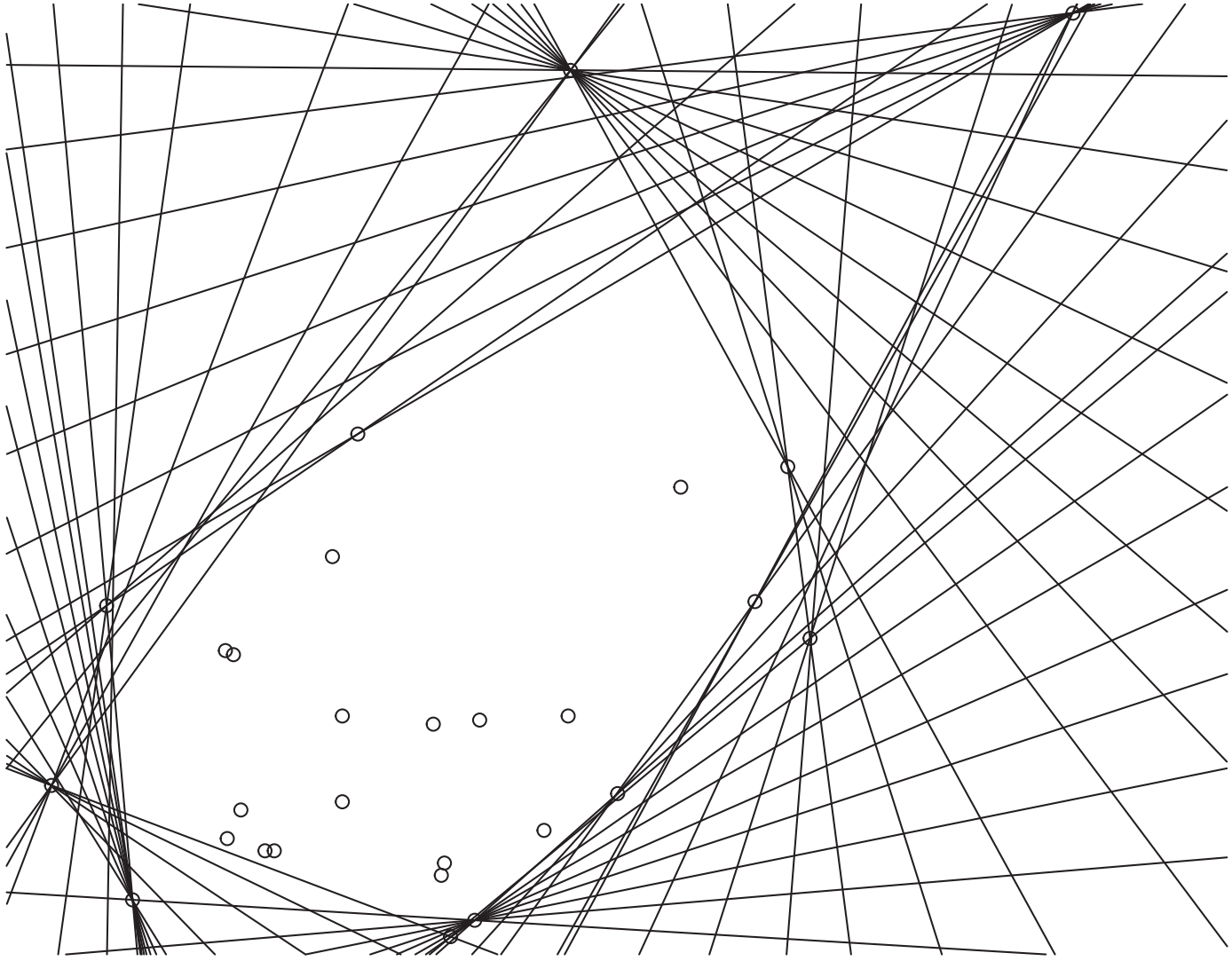
Directional quantiles (illustration)



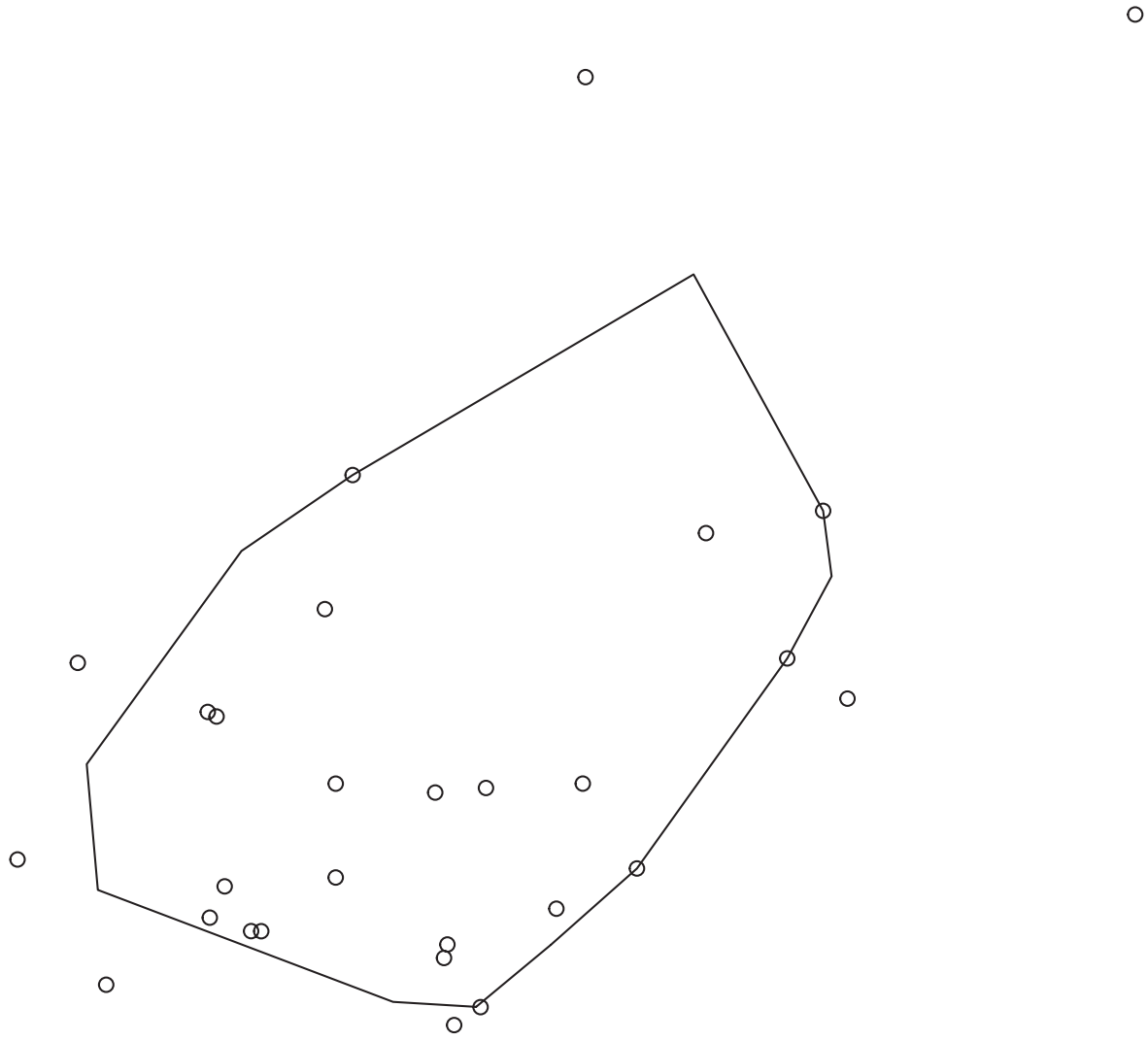
Directional quantiles (illustration)



Directional quantiles (illustration)



Directional quantiles (illustration)



Multiple-output regression quantiles

Hallin, Paindaveine, Šiman (2010); Paindaveine, Šiman (2011)

Regress $Y \in \mathbb{R}^m$ on $X = (1, W^T)^T \in \mathbb{R}^p$.

For a sample $(\boldsymbol{X}, \boldsymbol{Y}) = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathbb{R}^p \times \mathbb{R}^m; i = 1, ..., n\}$, for $\alpha \in (0, 1)$, and for $\boldsymbol{u} \in S^{m-1}$ a $(\alpha\boldsymbol{u})$-quantile positive halfspace is any

$$H^{(n)+}_{\alpha\boldsymbol{u}} := \{(\boldsymbol{w}^T, \boldsymbol{y}^T)^T \in \mathbb{R}^{p-1} \times \mathbb{R}^m : \hat{\boldsymbol{b}}^T_{\alpha\boldsymbol{u}}\boldsymbol{y} - \hat{\boldsymbol{a}}^T_{\alpha\boldsymbol{u}}(1, \boldsymbol{w}^T)^T \geq 0\} \text{ with}$$

$$(\hat{\boldsymbol{a}}^T_{\text{HP\v{S}};\alpha\boldsymbol{u}}, \hat{\boldsymbol{b}}^T_{\text{HP\v{S}};\alpha\boldsymbol{u}})^T = \arg\min \sum_{i=1}^{n} \rho_\alpha(\boldsymbol{b}^T\boldsymbol{y}_i - \boldsymbol{a}^T\boldsymbol{x}_i) \text{ subject to } \boldsymbol{u}^T\boldsymbol{b} = 1,$$

or

$$(\hat{\boldsymbol{a}}^T_{\text{proj};\alpha\boldsymbol{u}}, \hat{\boldsymbol{b}}^T_{\text{proj};\alpha\boldsymbol{u}})^T = \arg\min \sum_{i=1}^{n} \rho_\alpha(\boldsymbol{b}^T\boldsymbol{y}_i - \boldsymbol{a}^T\boldsymbol{x}_i) \text{ subject to } \boldsymbol{u} = \boldsymbol{b},$$

where $\rho_\alpha(x) = x\big(\alpha - I(x < 0)\big)$ is the α-quantile check function.

Multiple-output regression quantiles (location)

Hallin, Paindaveine, Šiman (2010); Paindaveine, Šiman (2011)

Regress $X \in \mathbb{R}^d$ on $1 \in \mathbb{R}$.

For a sample $\boldsymbol{X} = \{\boldsymbol{x}_i \in \mathbb{R}^d; i = 1, ..., n\}$, for $\alpha \in (0, 1)$, and for $\boldsymbol{u} \in S^{d-1}$ a $(\alpha\boldsymbol{u})$-quantile positive halfspace is any

$$H^{(n)+}_{\alpha\boldsymbol{u}} := \{\boldsymbol{x}^T \in \mathbb{R}^d \ : \ \hat{\boldsymbol{b}}^T_{\alpha\boldsymbol{u}}\boldsymbol{x} - \hat{a}_{\alpha\boldsymbol{u}} \geq 0\} \text{ with}$$

$$(\hat{a}_{\text{HP}\check{\text{S}};\alpha\boldsymbol{u}}, \hat{\boldsymbol{b}}^T_{\text{HP}\check{\text{S}};\alpha\boldsymbol{u}})^T = \arg\min \sum_{i=1}^{n} \rho_\alpha(\boldsymbol{b}^T\boldsymbol{x}_i - a) \text{ subject to } \boldsymbol{u}^T\boldsymbol{b} = 1,$$

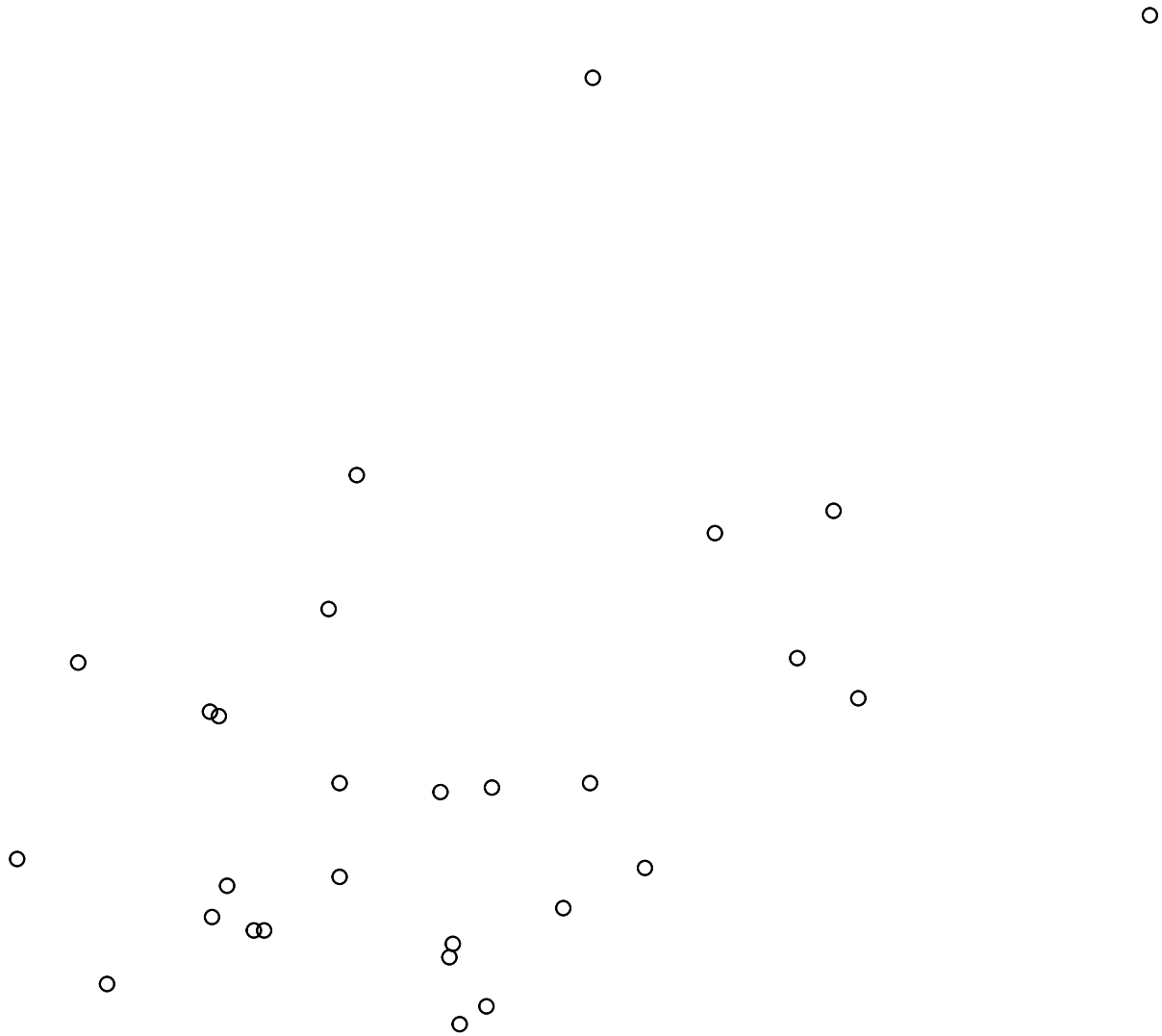
or

$$(\hat{a}_{\text{proj};\alpha\boldsymbol{u}}, \hat{\boldsymbol{b}}^T_{\text{proj};\alpha\boldsymbol{u}})^T = \arg\min \sum_{i=1}^{n} \rho_\alpha(\boldsymbol{b}^T\boldsymbol{x}_i - a) \text{ subject to } \boldsymbol{u} = \boldsymbol{b},$$

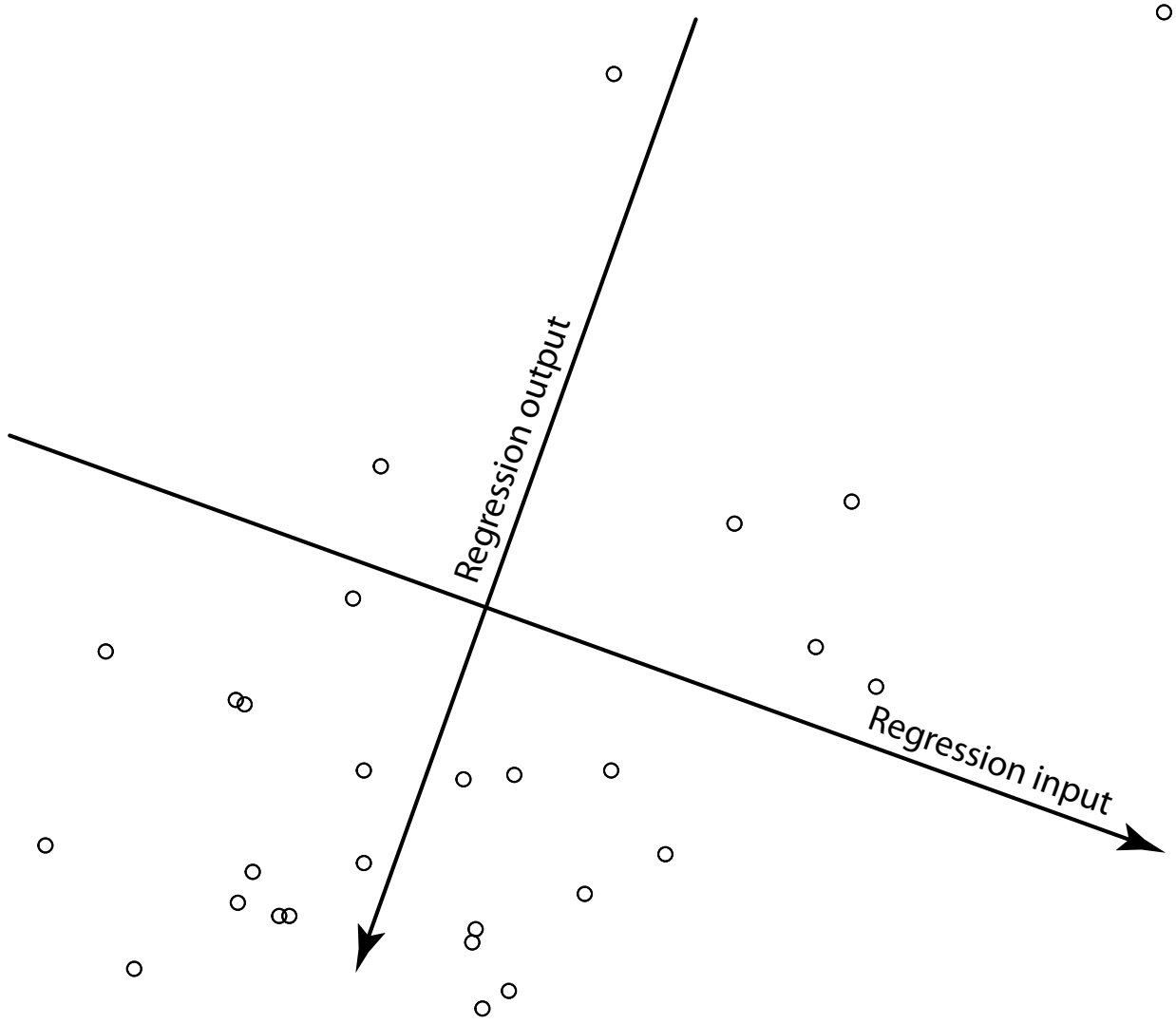
where $\rho_\alpha(x) = x\big(\alpha - I(x < 0)\big)$ is the α-quantile check function.

$$\text{Then } D^{Tuk}_\alpha(\boldsymbol{X}) = R_\alpha(\boldsymbol{X}) = \bigcap_{\boldsymbol{u} \in S^{d-1}} \{H^{(n)+}_{\alpha\boldsymbol{u}}\}.$$

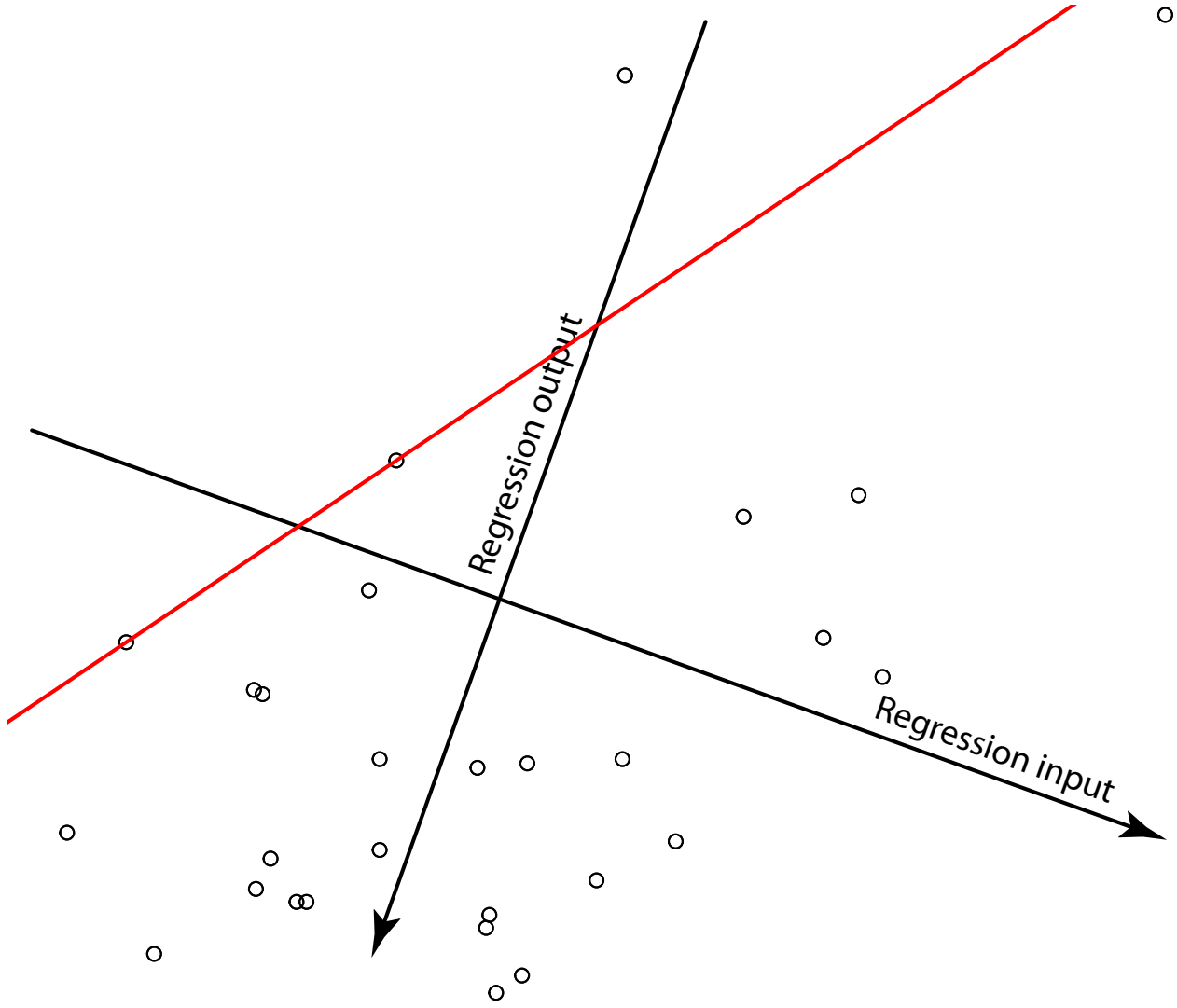
Multiple-output regression quantiles (illustration)



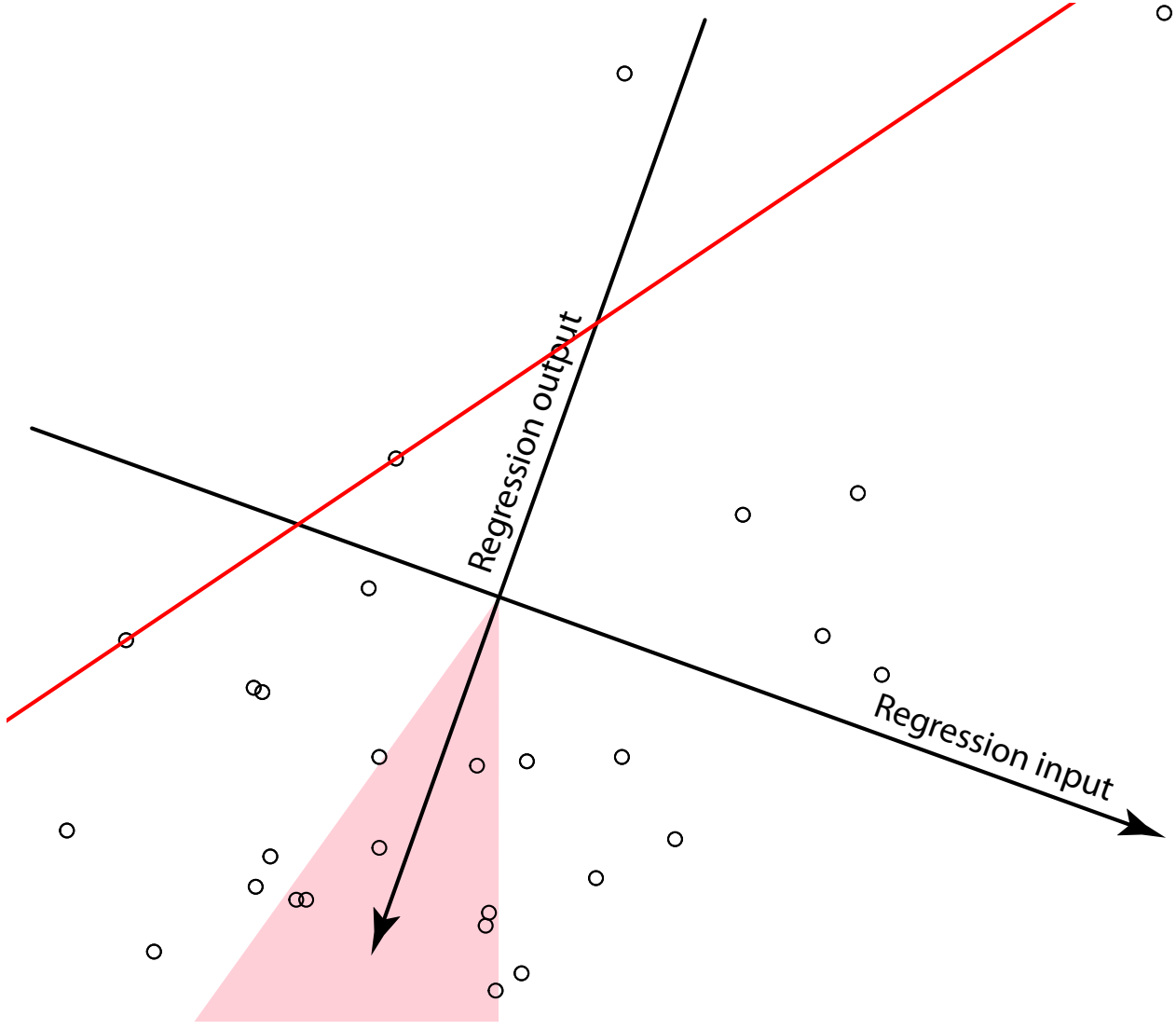
Multiple-output regression quantiles (illustration)



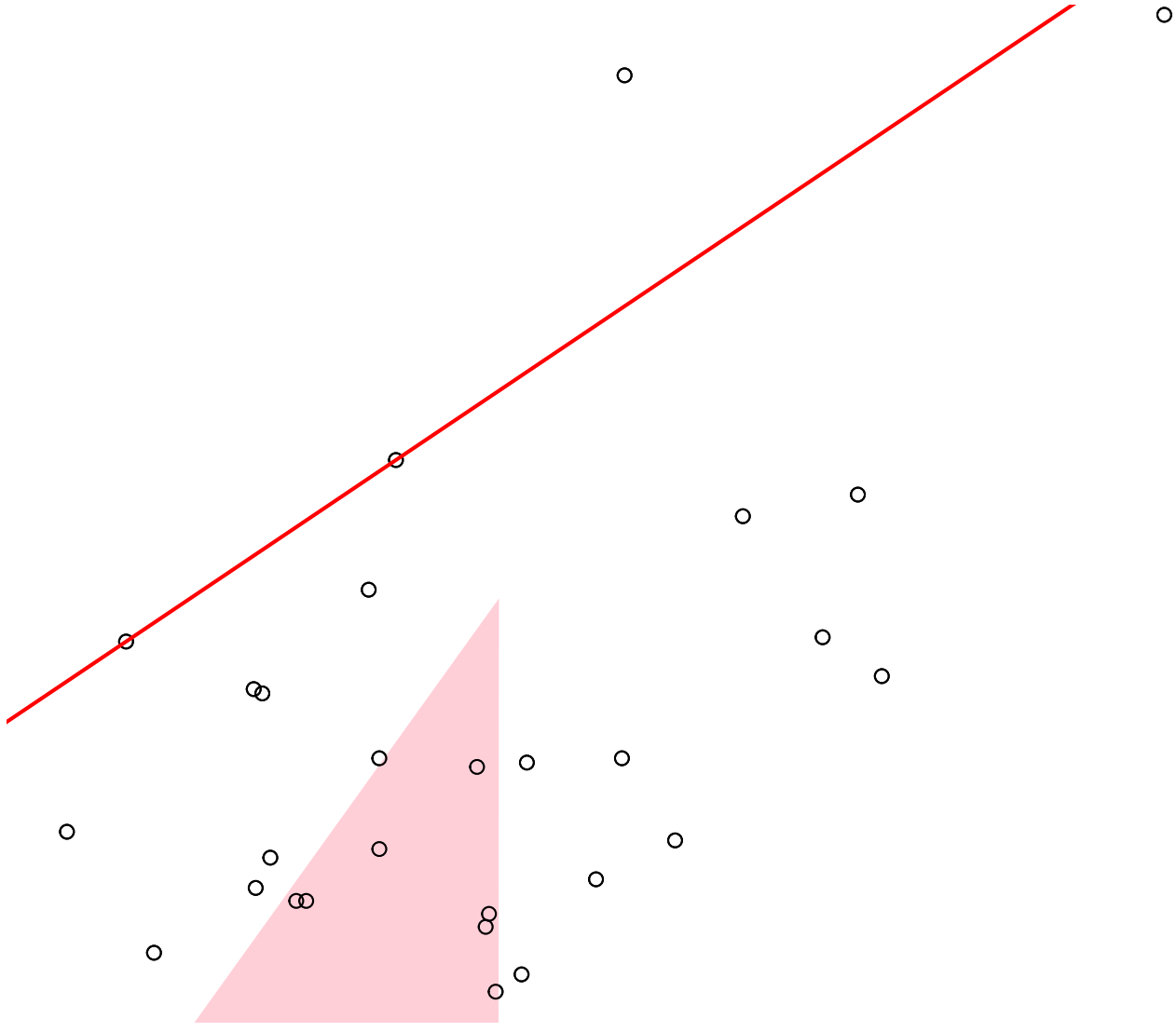
Multiple-output regression quantiles (illustration)



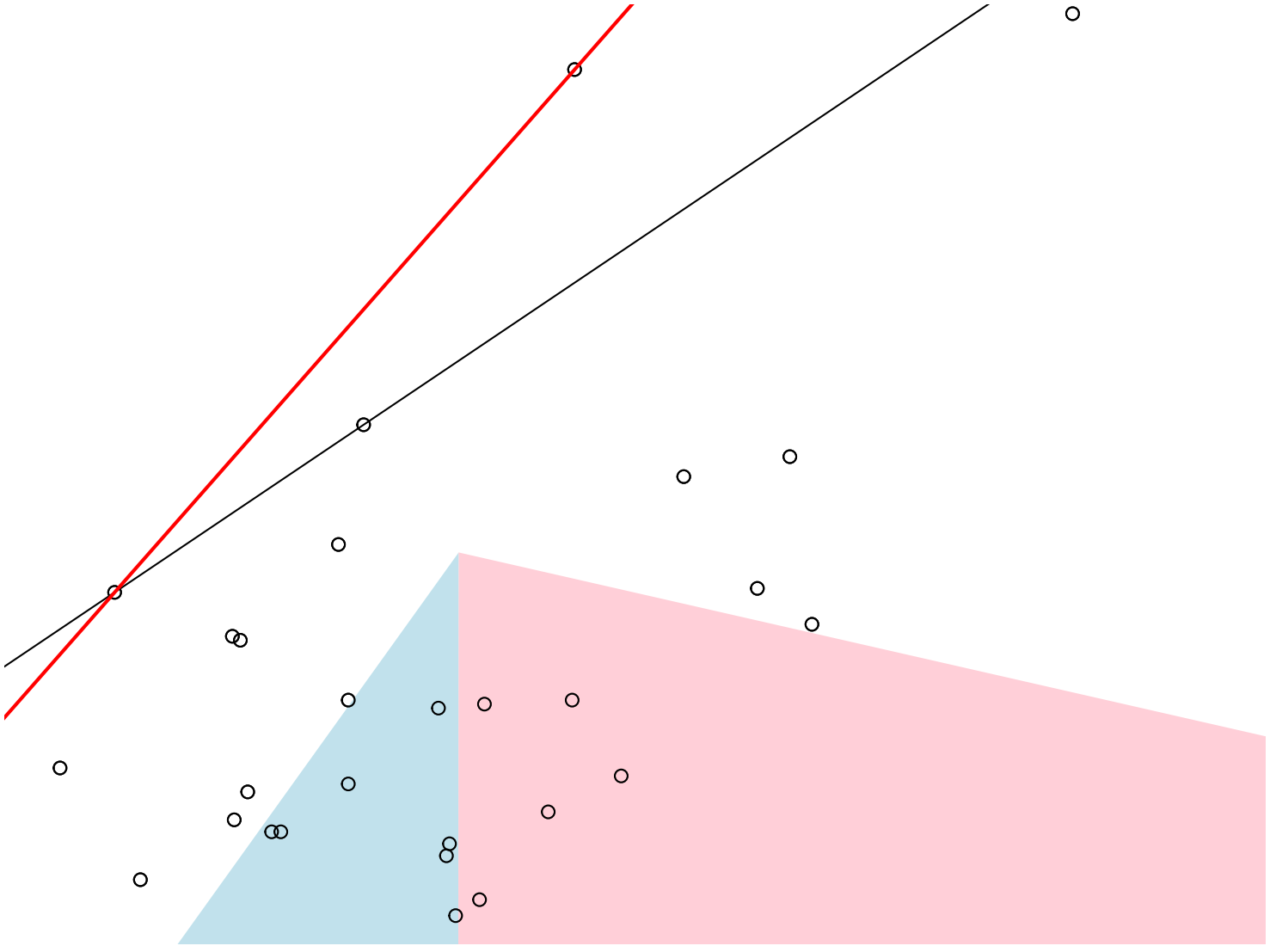
Multiple-output regression quantiles (illustration)



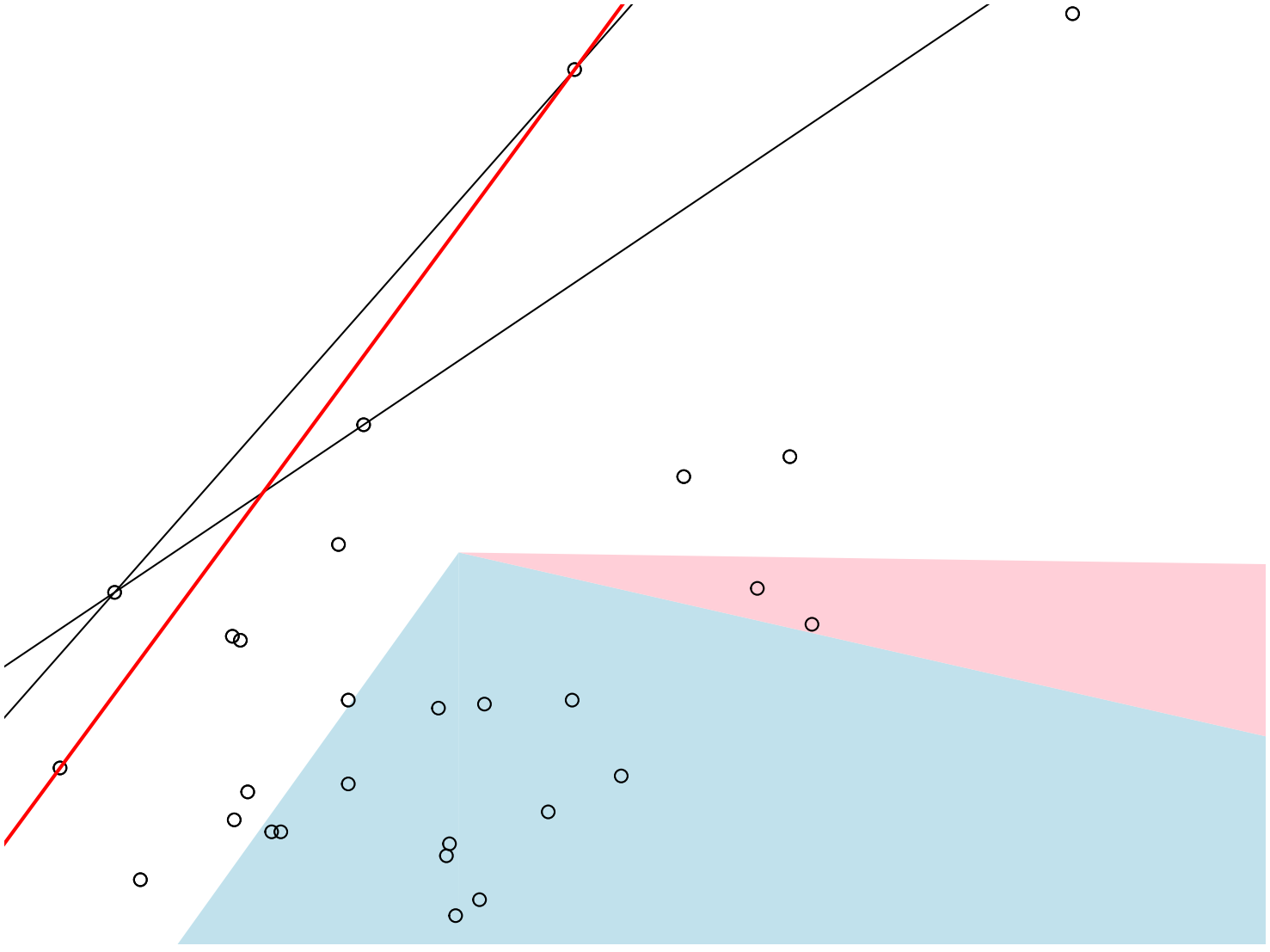
Multiple-output regression quantiles (illustration)



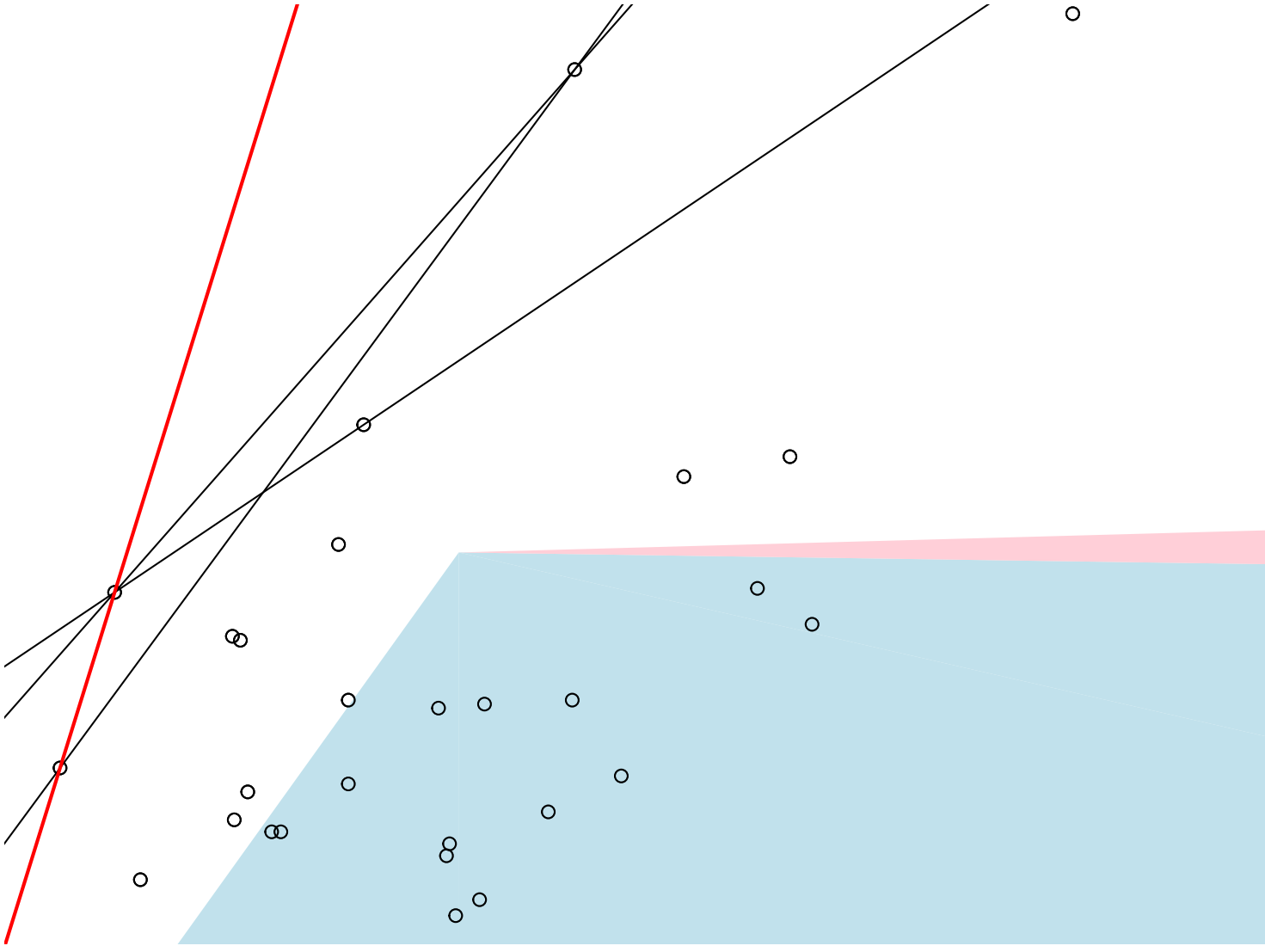
Multiple-output regression quantiles (illustration)



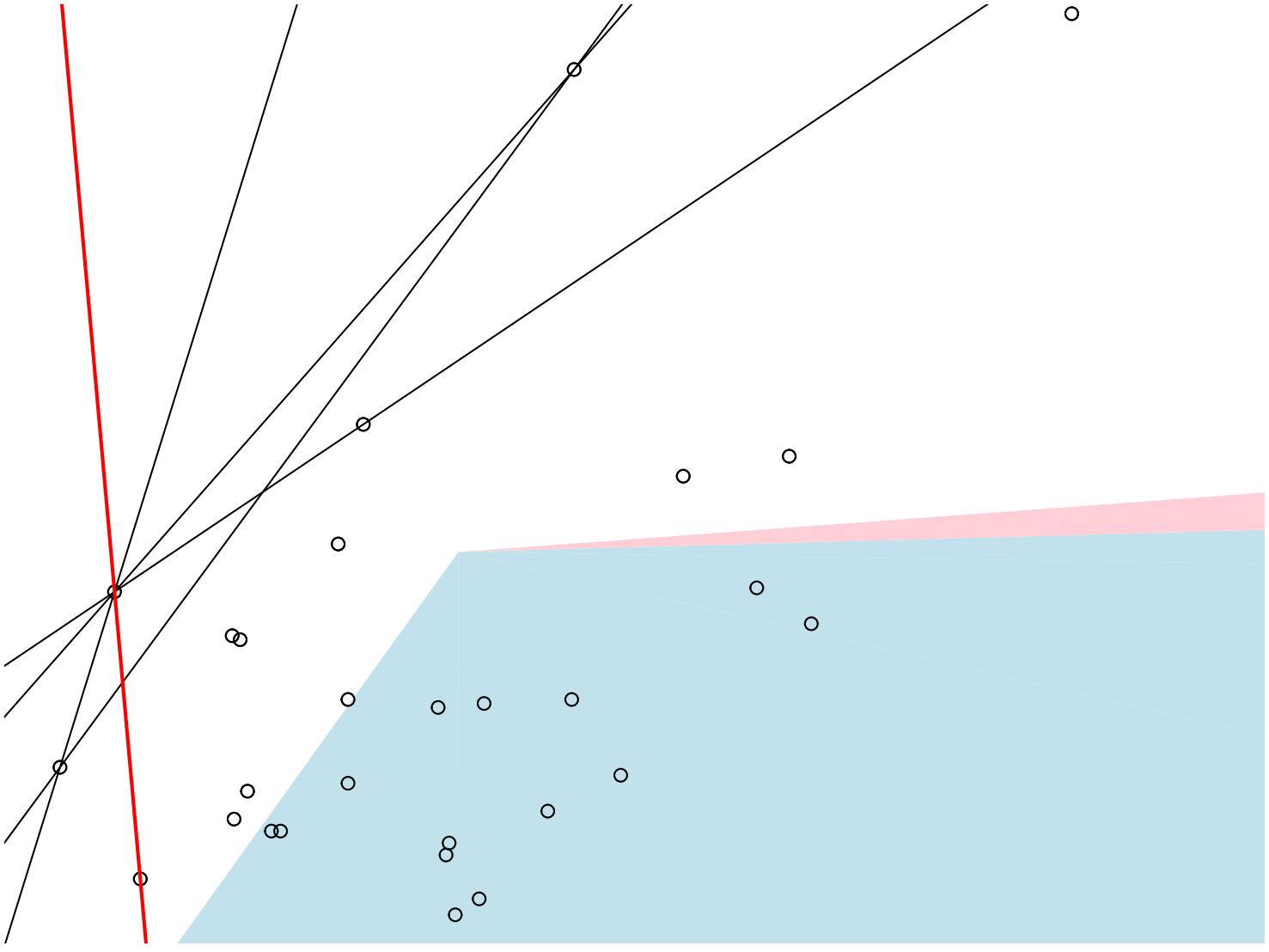
Multiple-output regression quantiles (illustration)



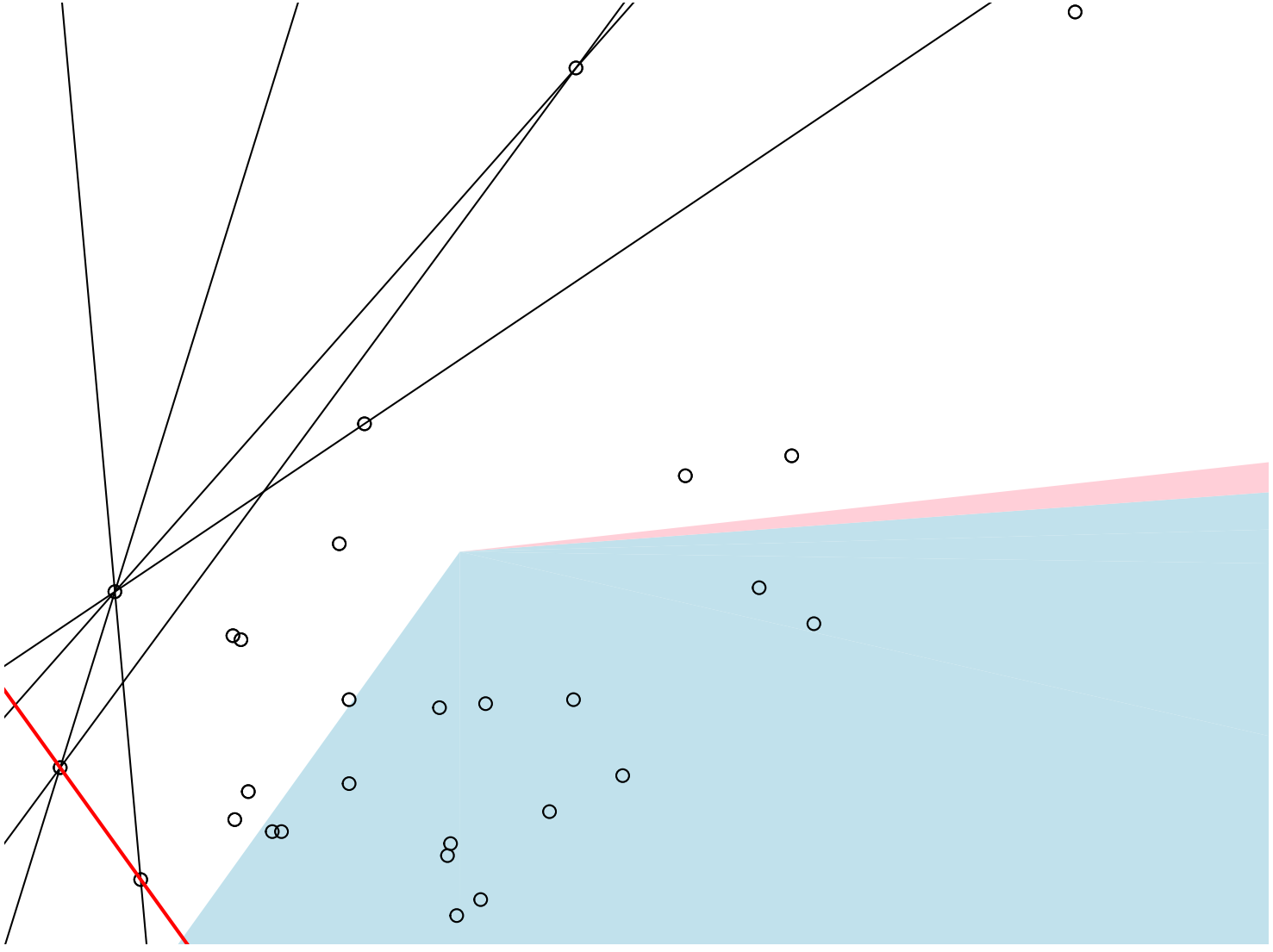
Multiple-output regression quantiles (illustration)



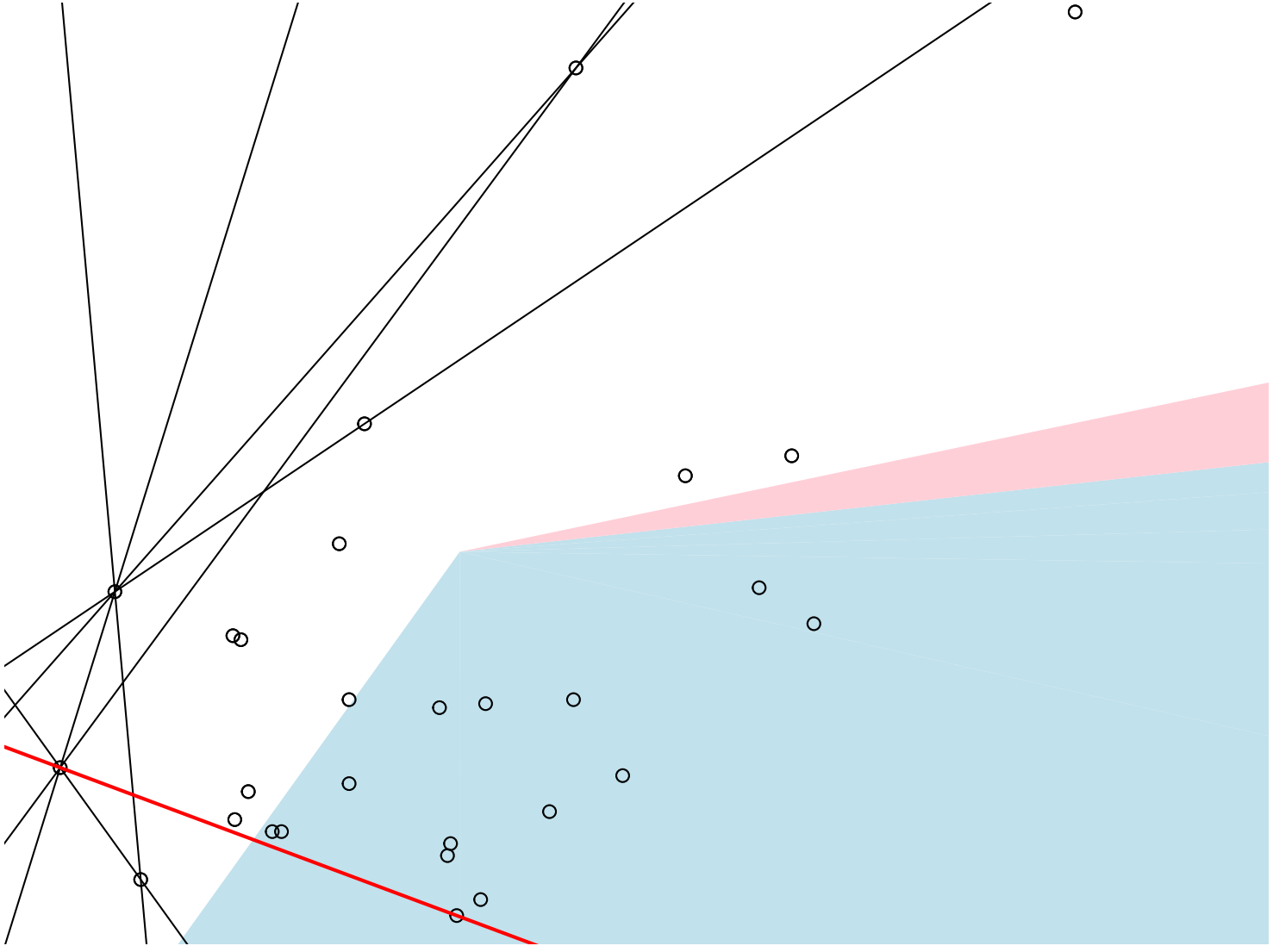
Multiple-output regression quantiles (illustration)



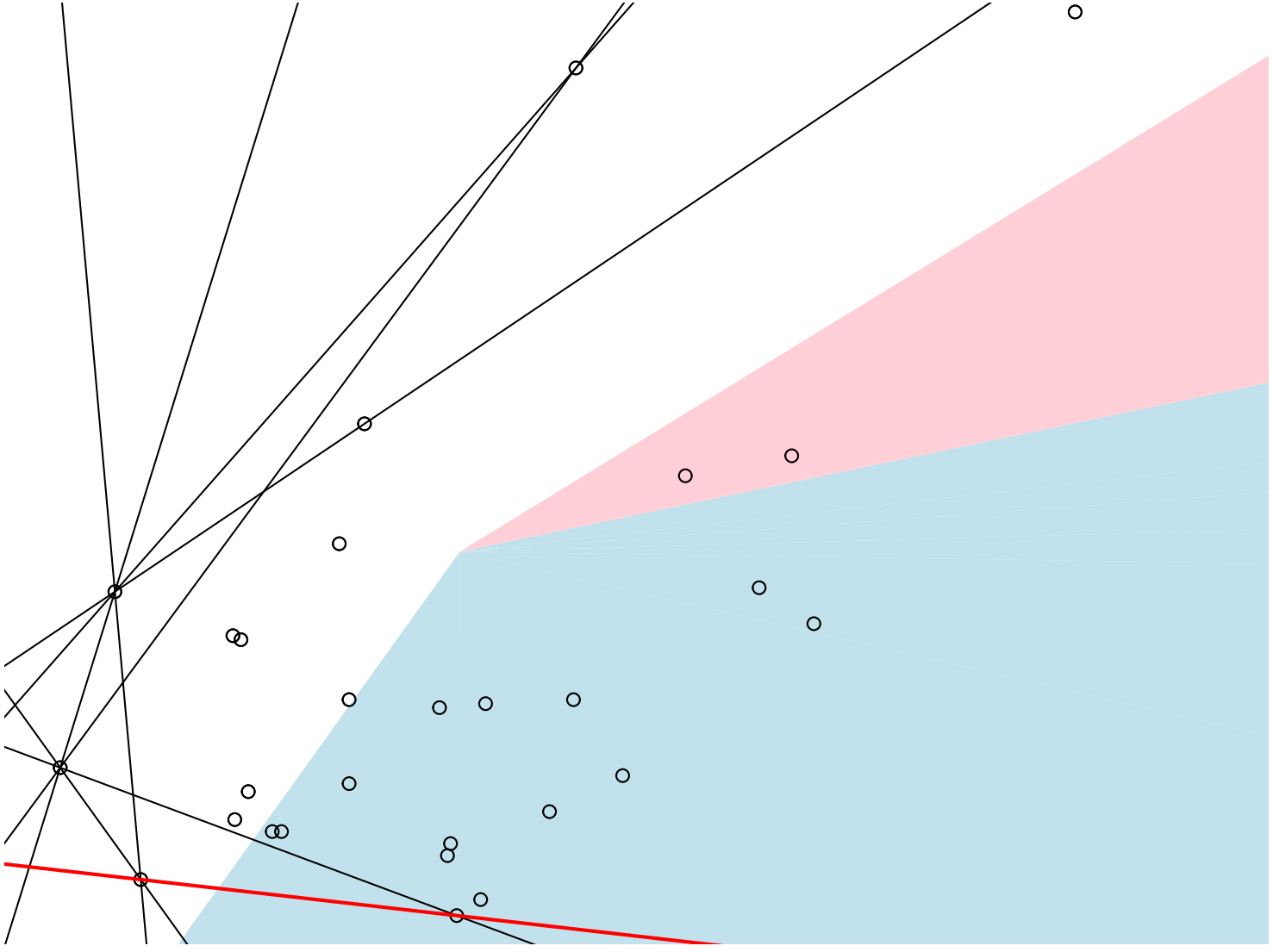
Multiple-output regression quantiles (illustration)



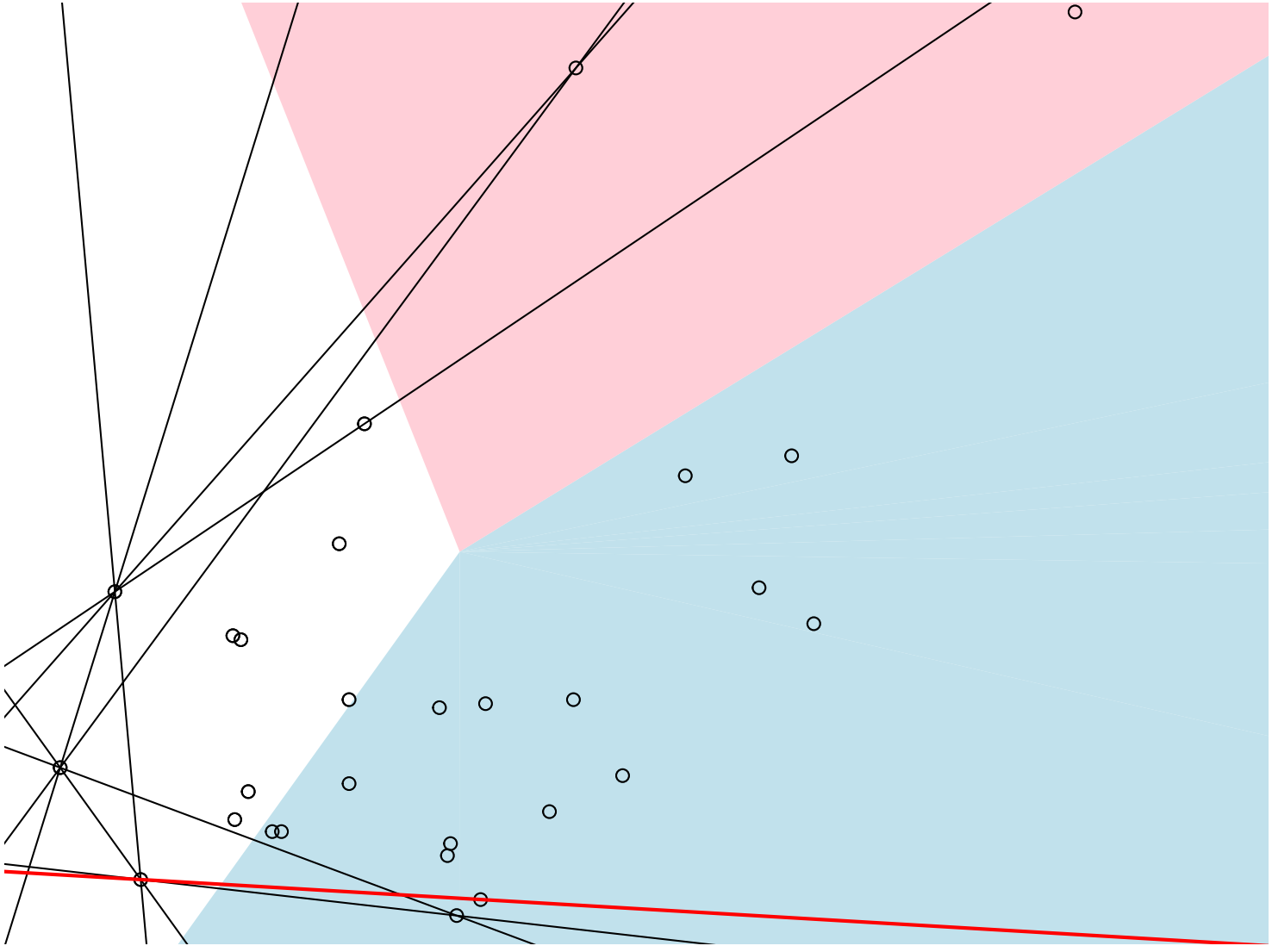
Multiple-output regression quantiles (illustration)



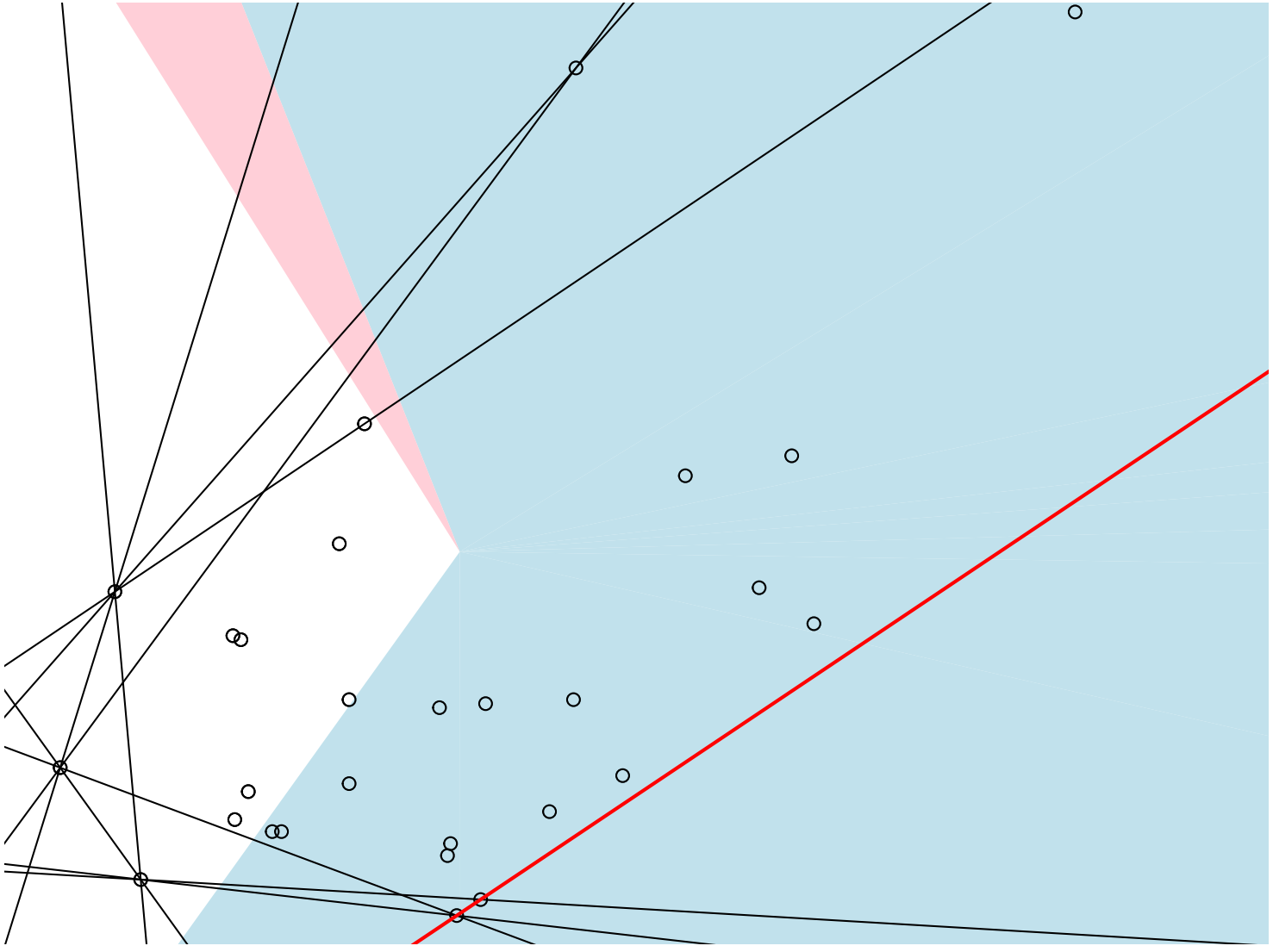
Multiple-output regression quantiles (illustration)



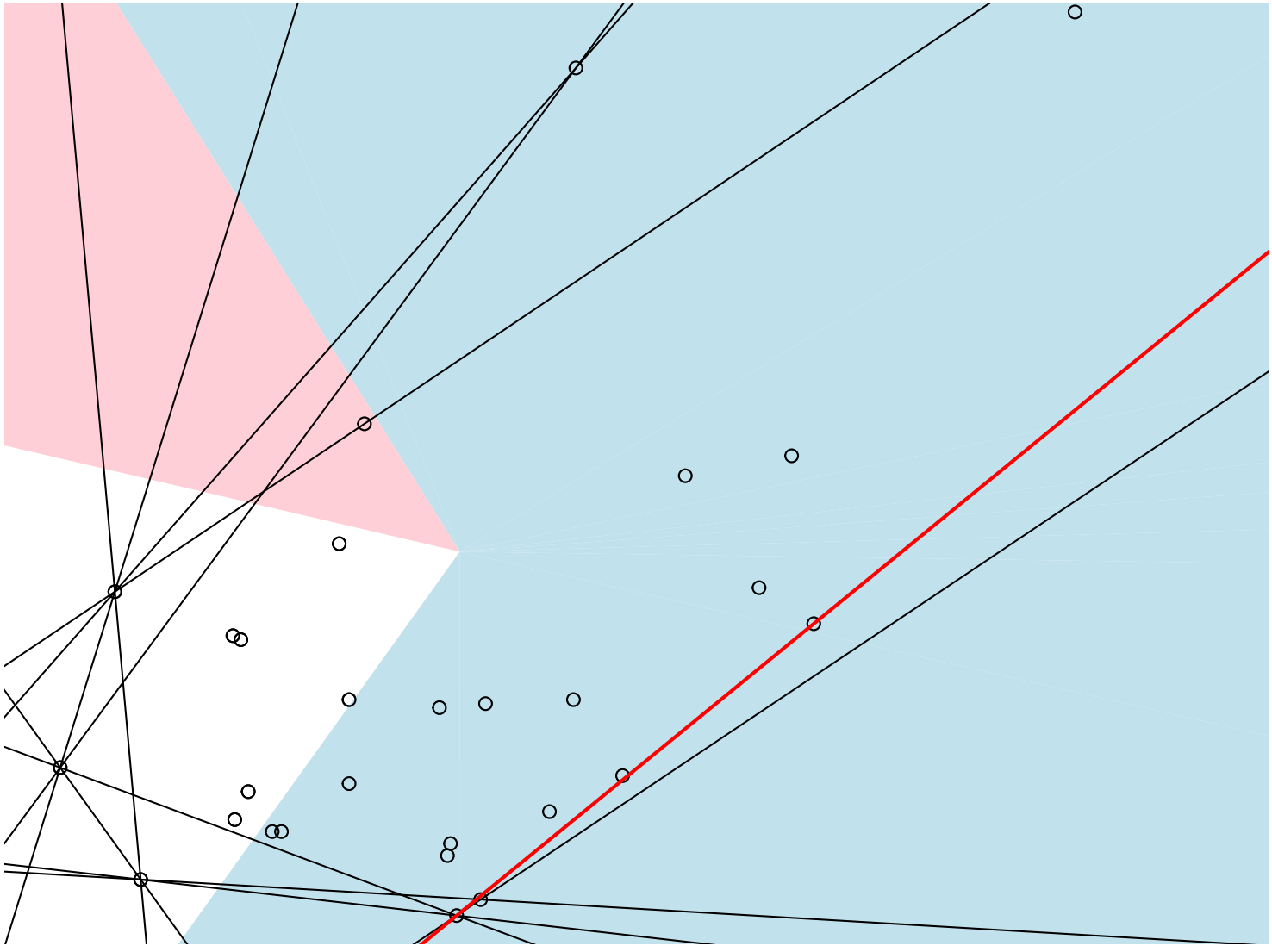
Multiple-output regression quantiles (illustration)



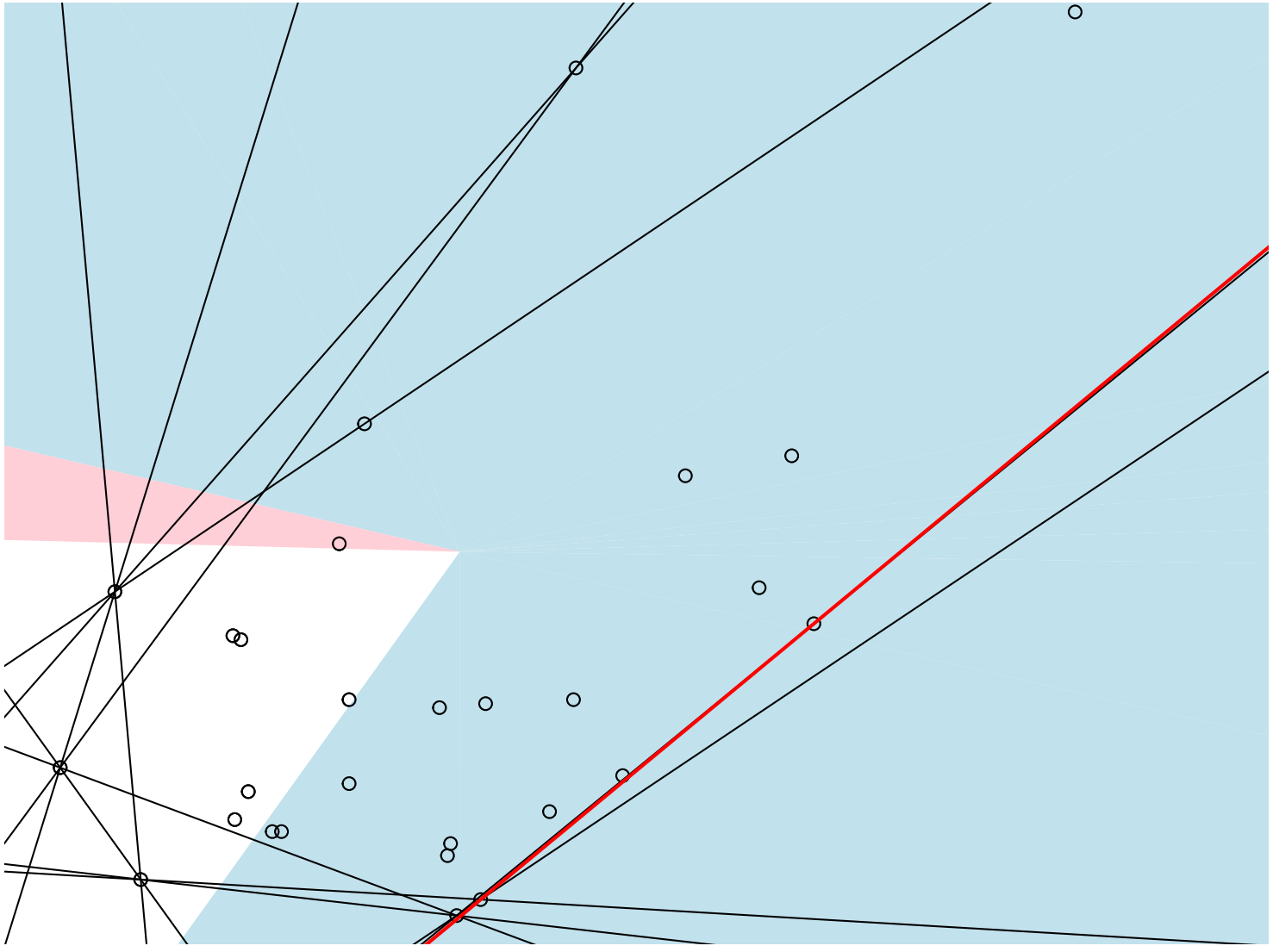
Multiple-output regression quantiles (illustration)



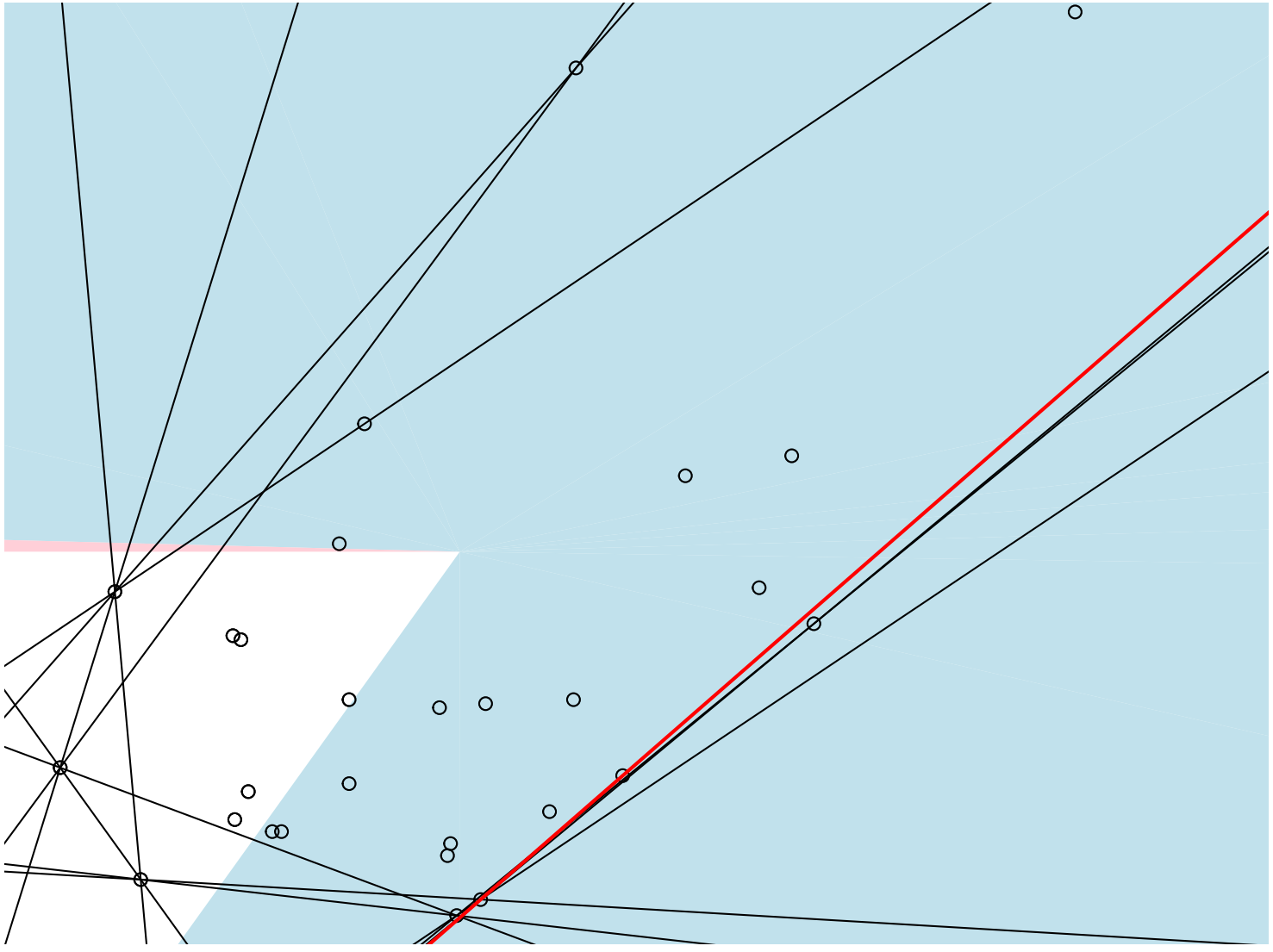
Multiple-output regression quantiles (illustration)



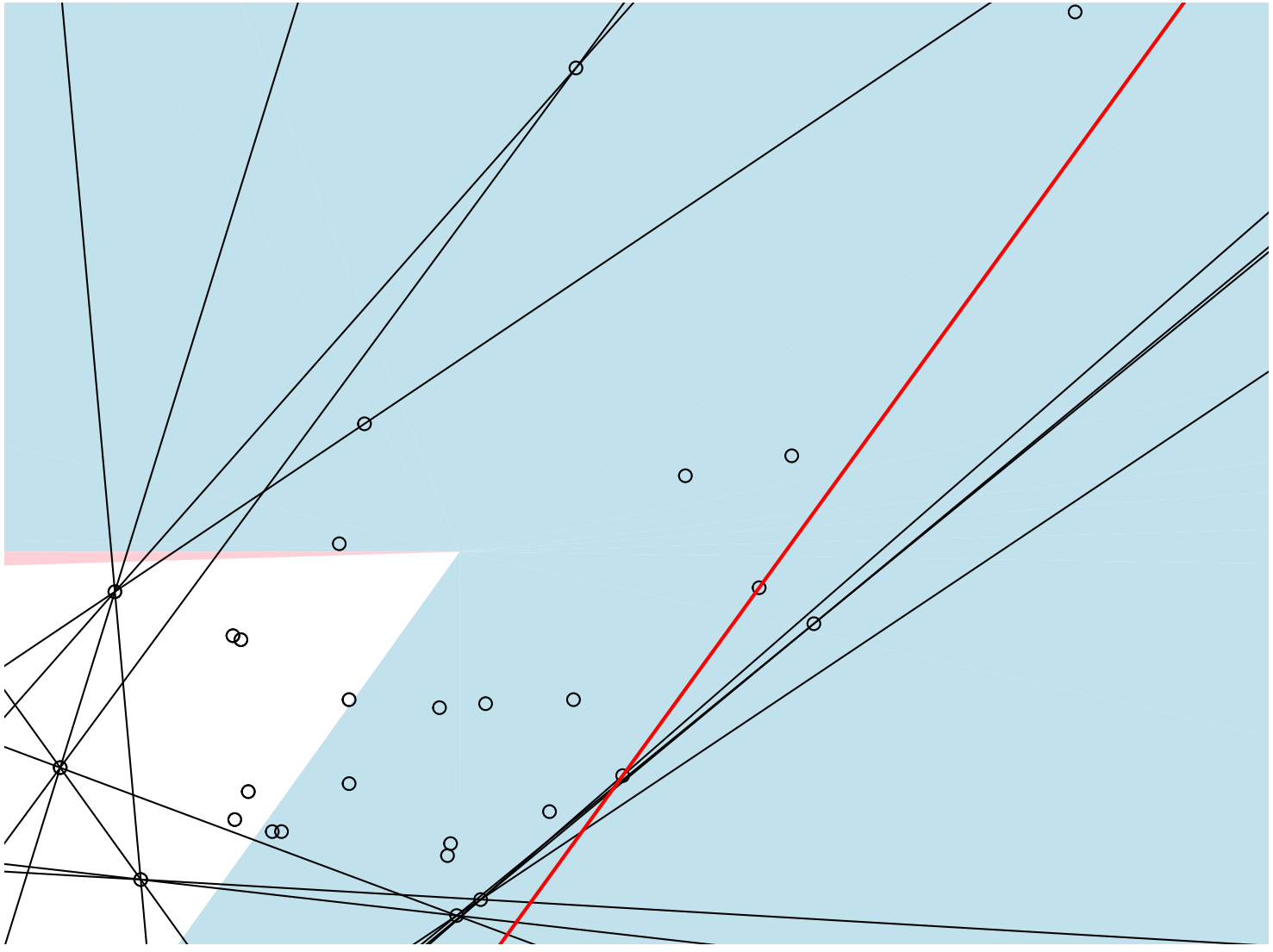
Multiple-output regression quantiles (illustration)



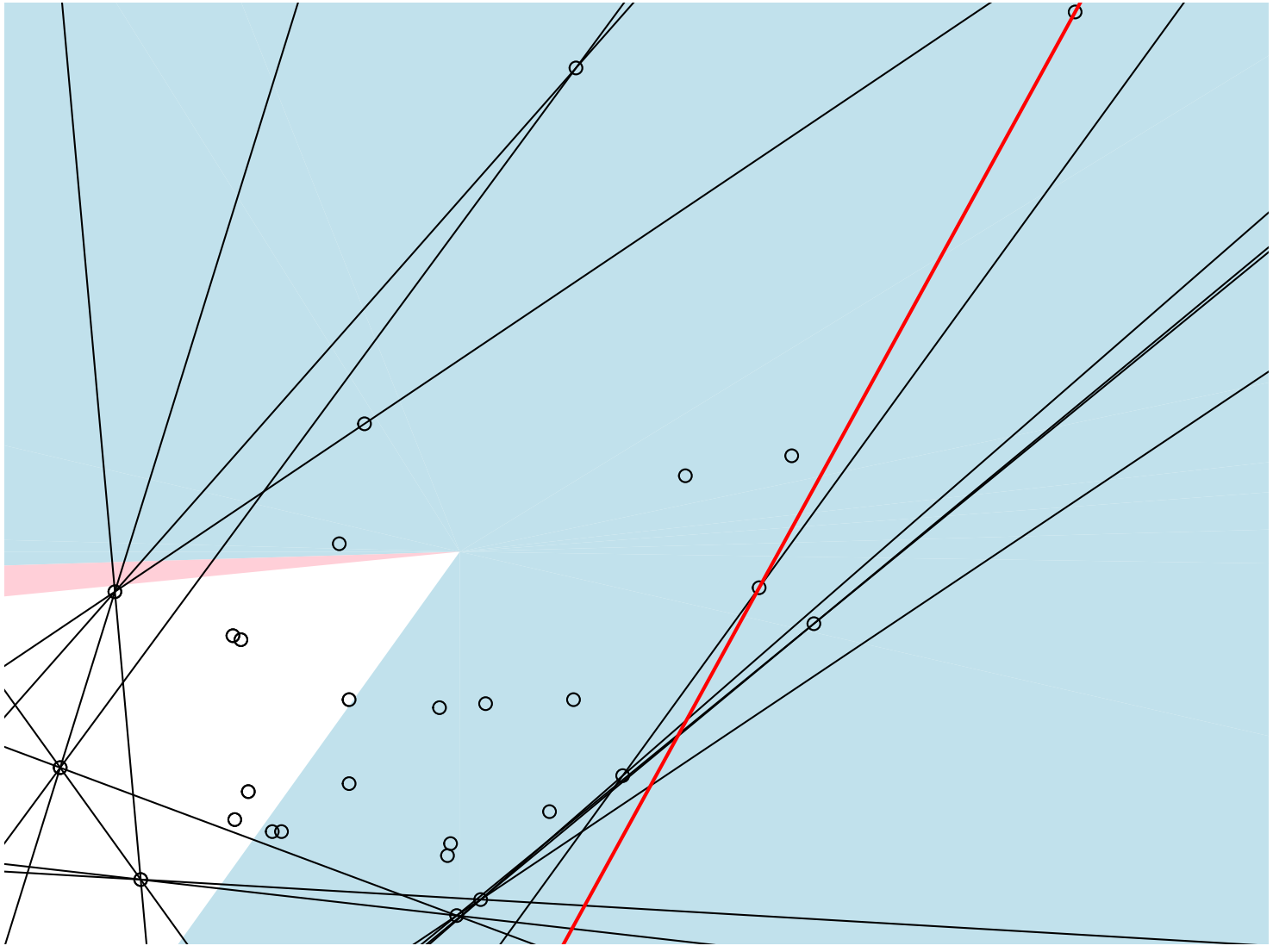
Multiple-output regression quantiles (illustration)



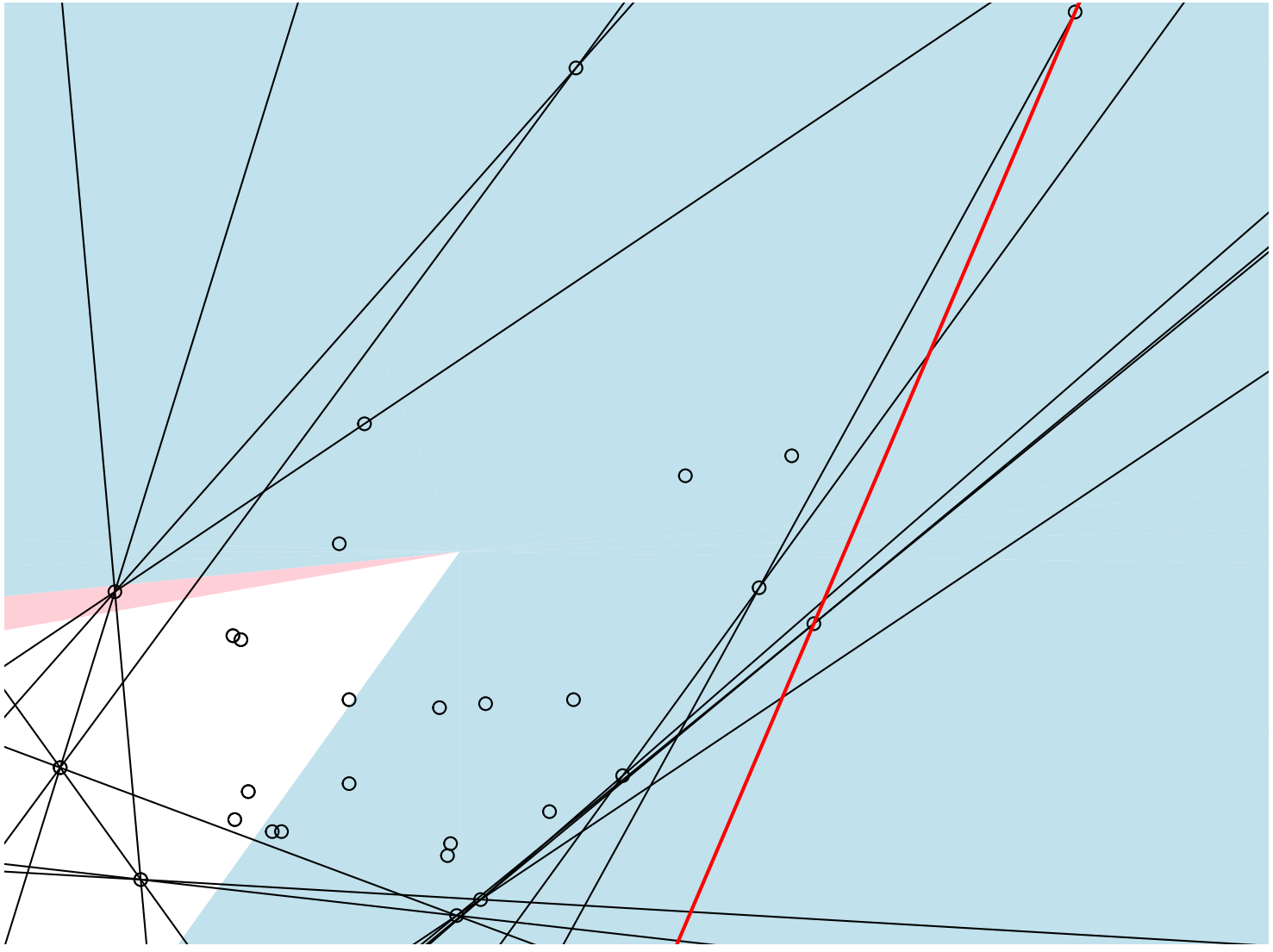
Multiple-output regression quantiles (illustration)



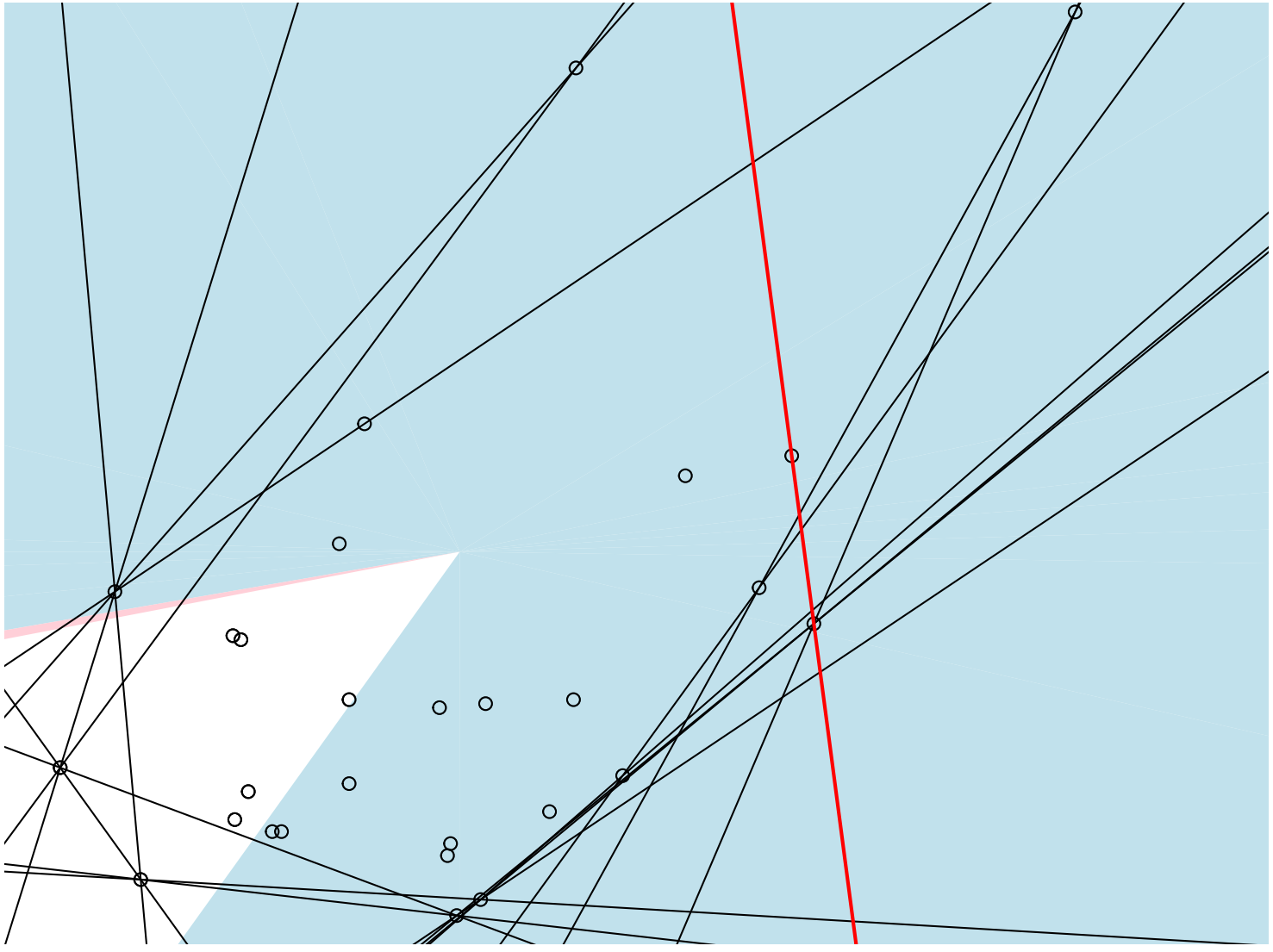
Multiple-output regression quantiles (illustration)



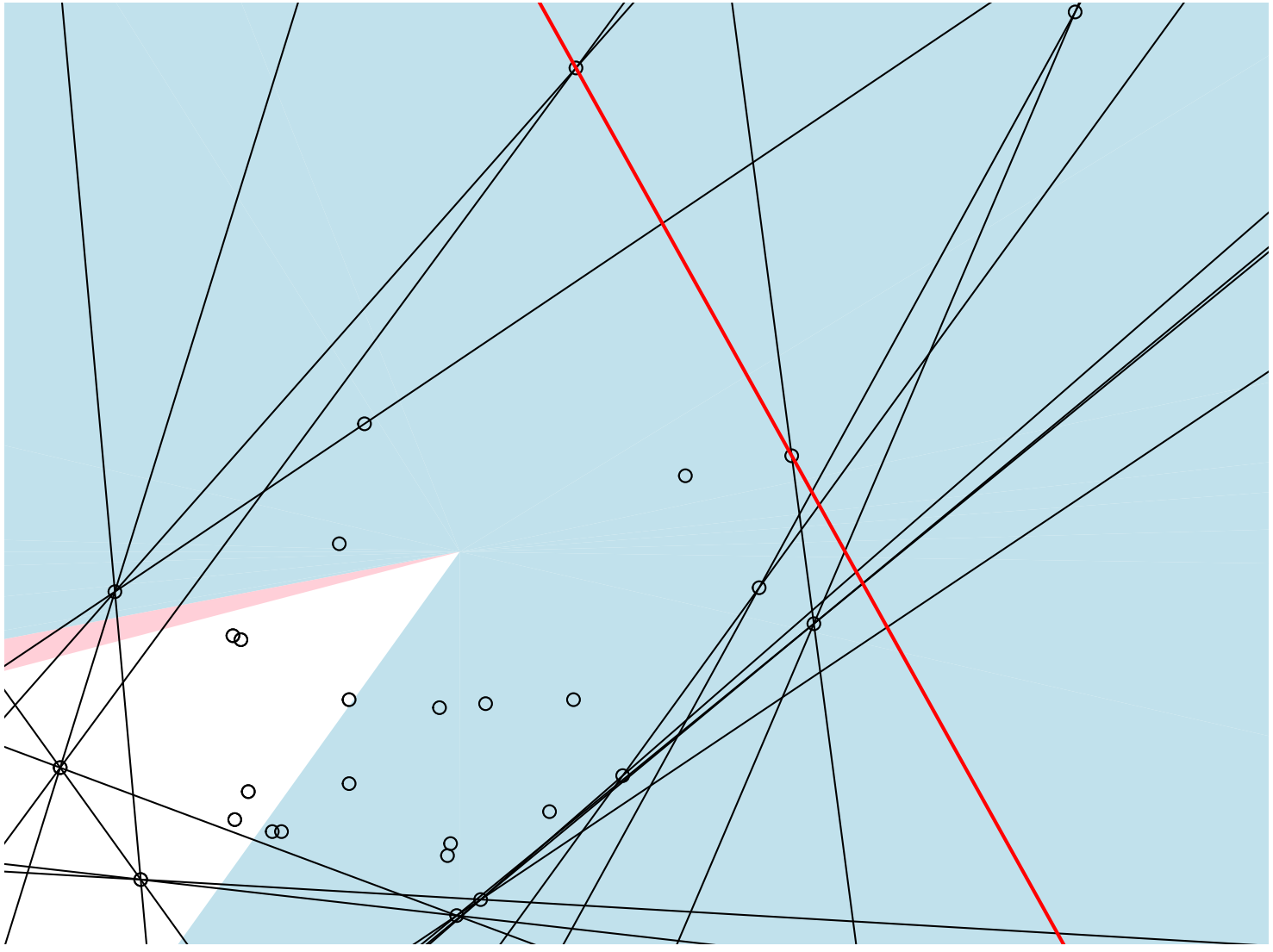
Multiple-output regression quantiles (illustration)



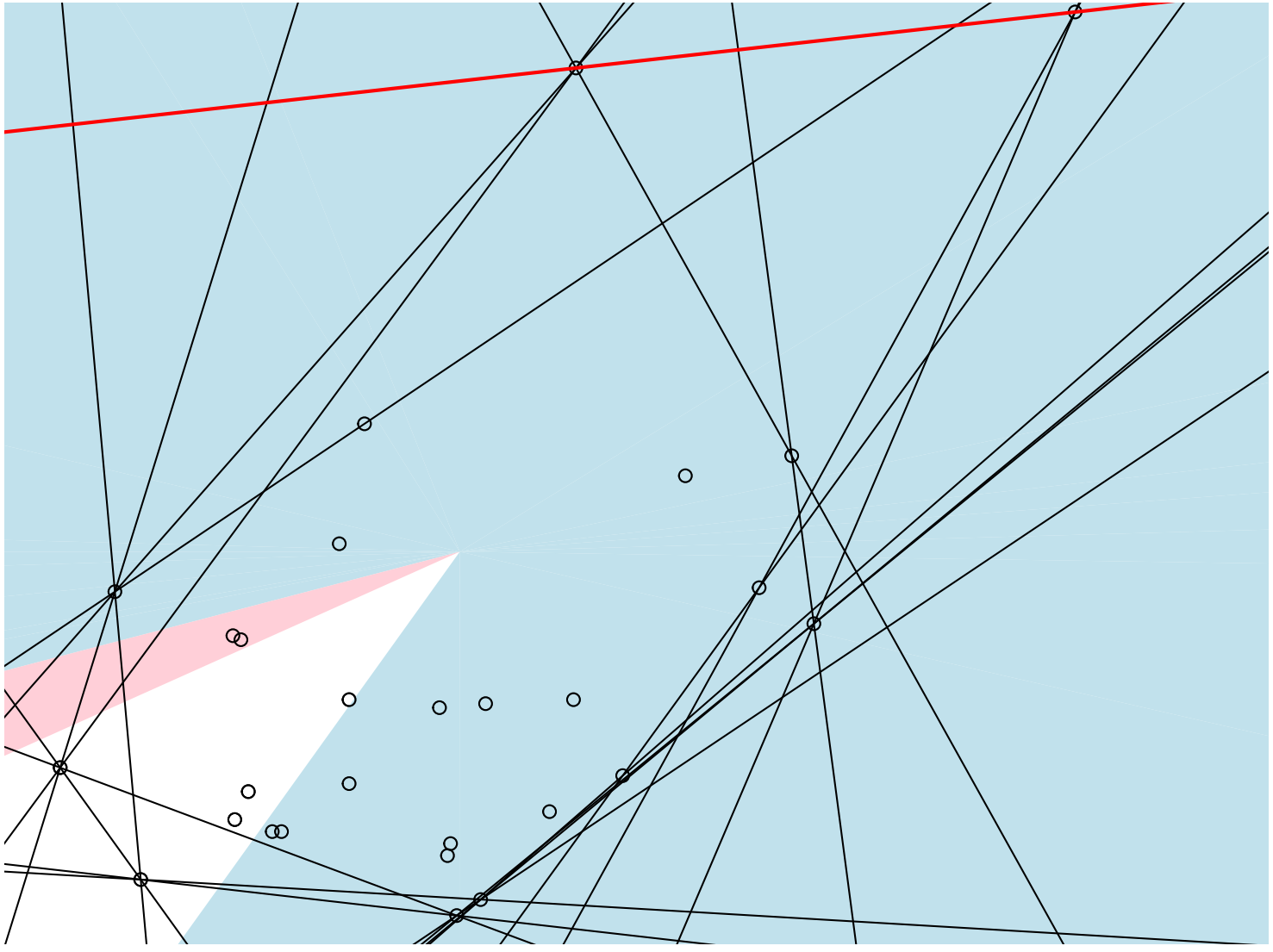
Multiple-output regression quantiles (illustration)



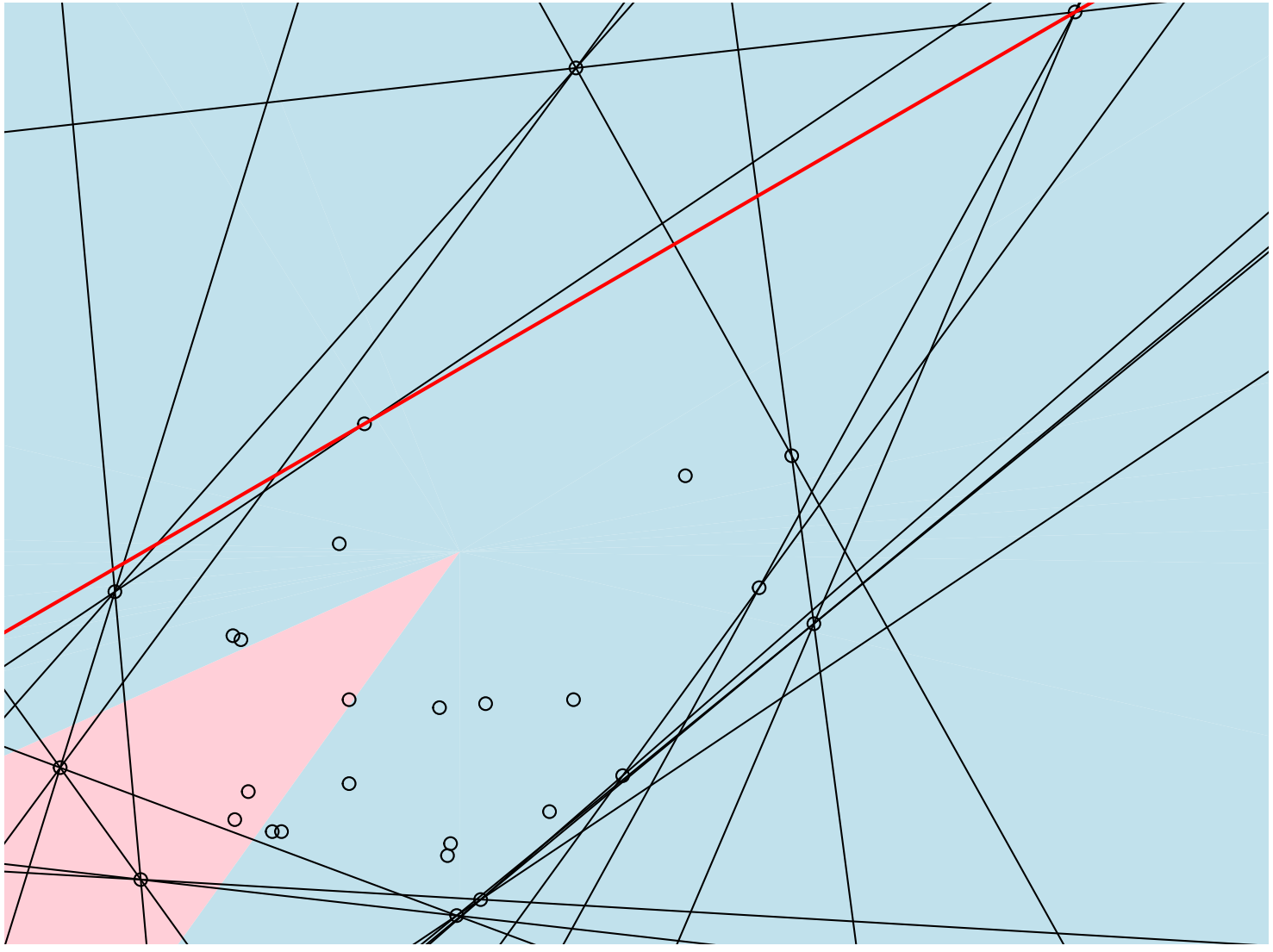
Multiple-output regression quantiles (illustration)



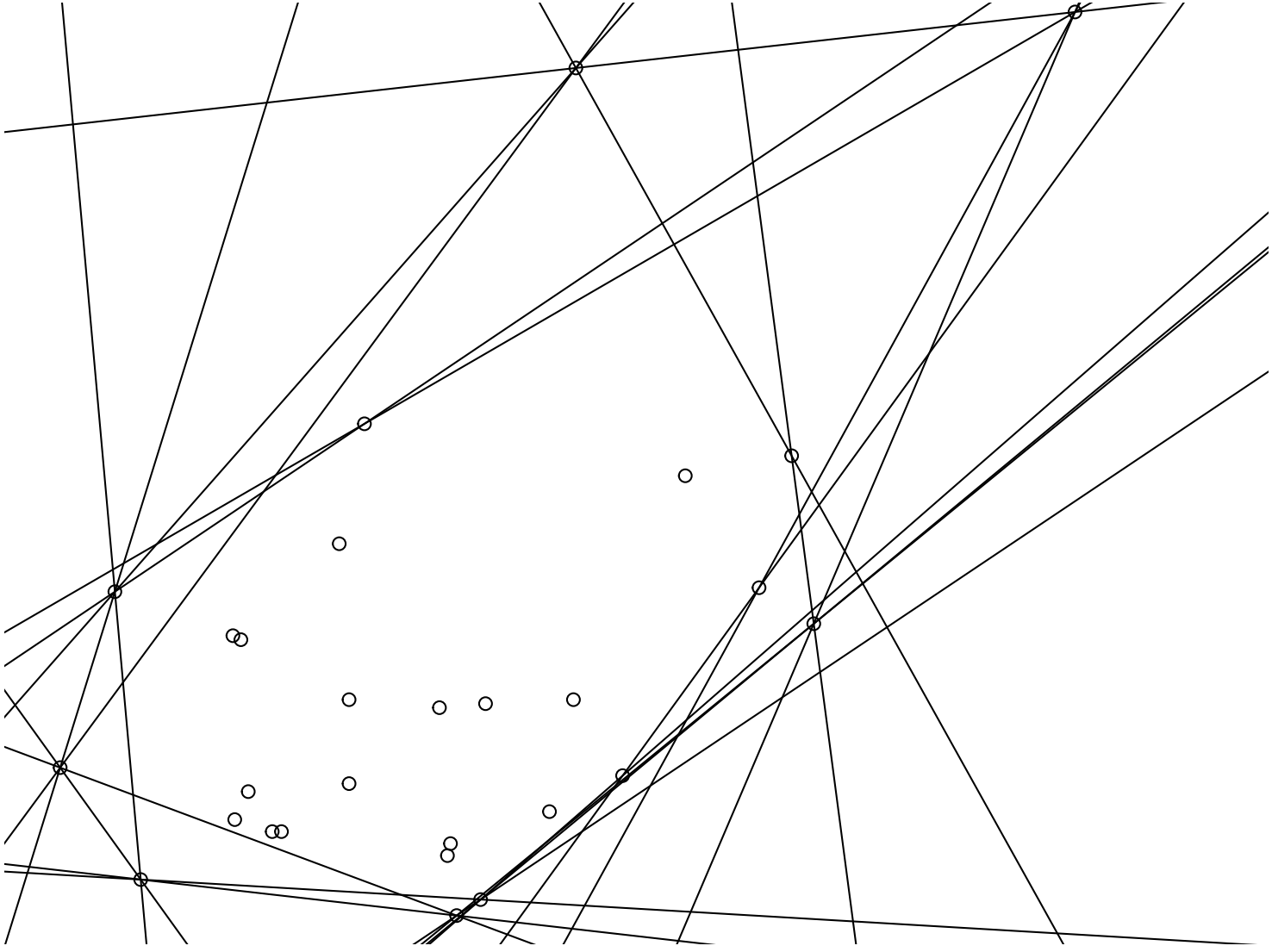
Multiple-output regression quantiles (illustration)



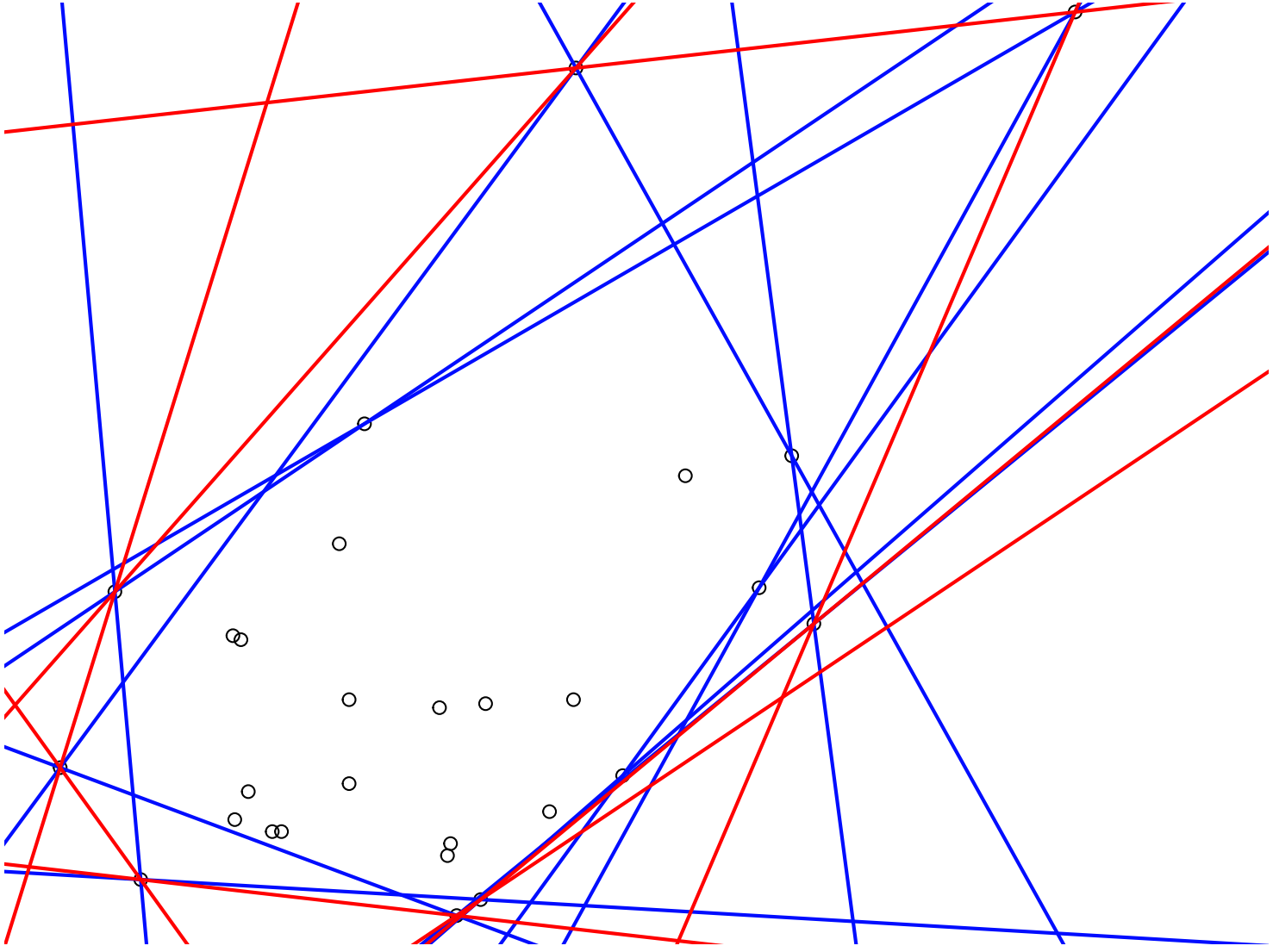
Multiple-output regression quantiles (illustration)



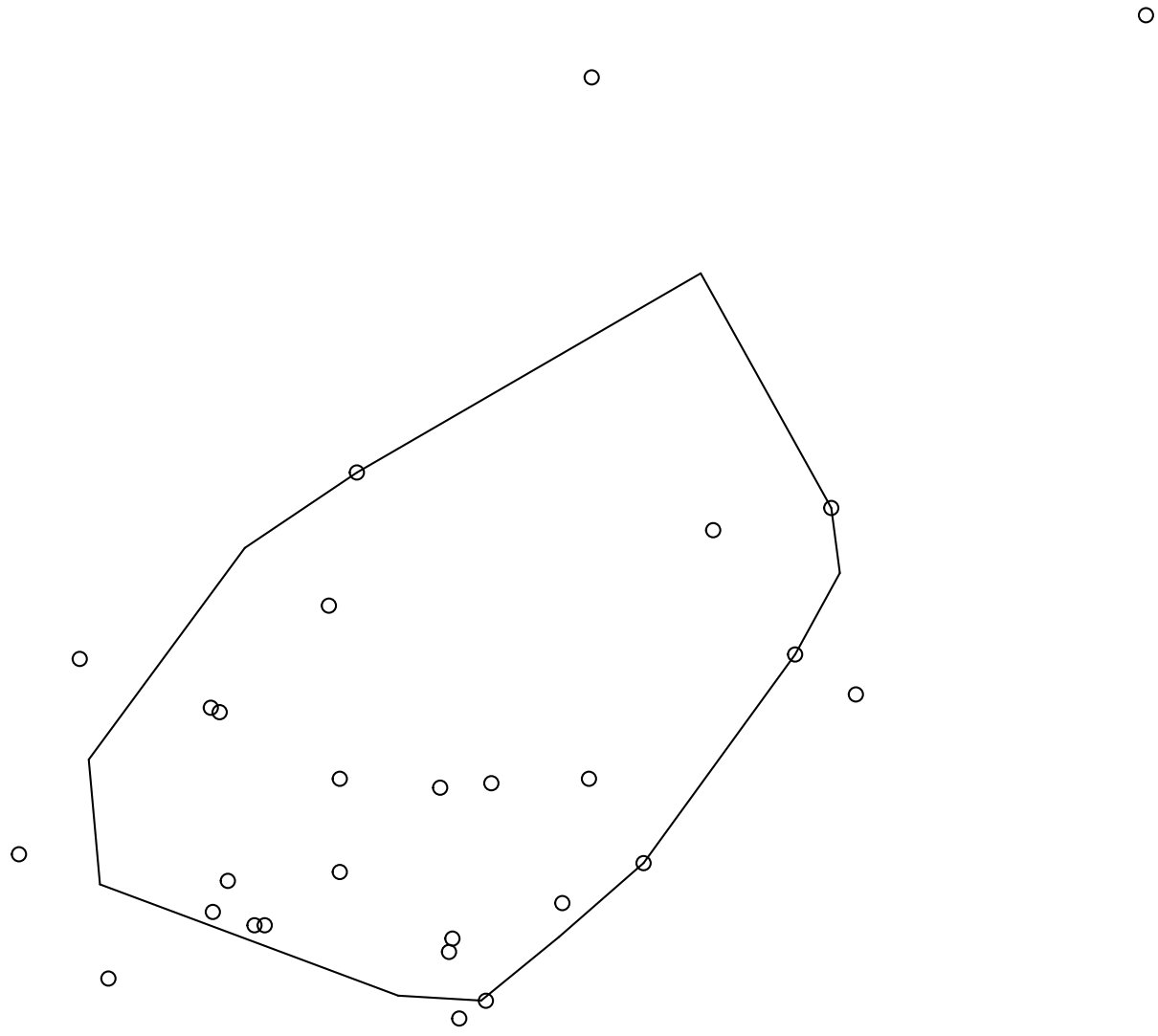
Multiple-output regression quantiles (illustration)



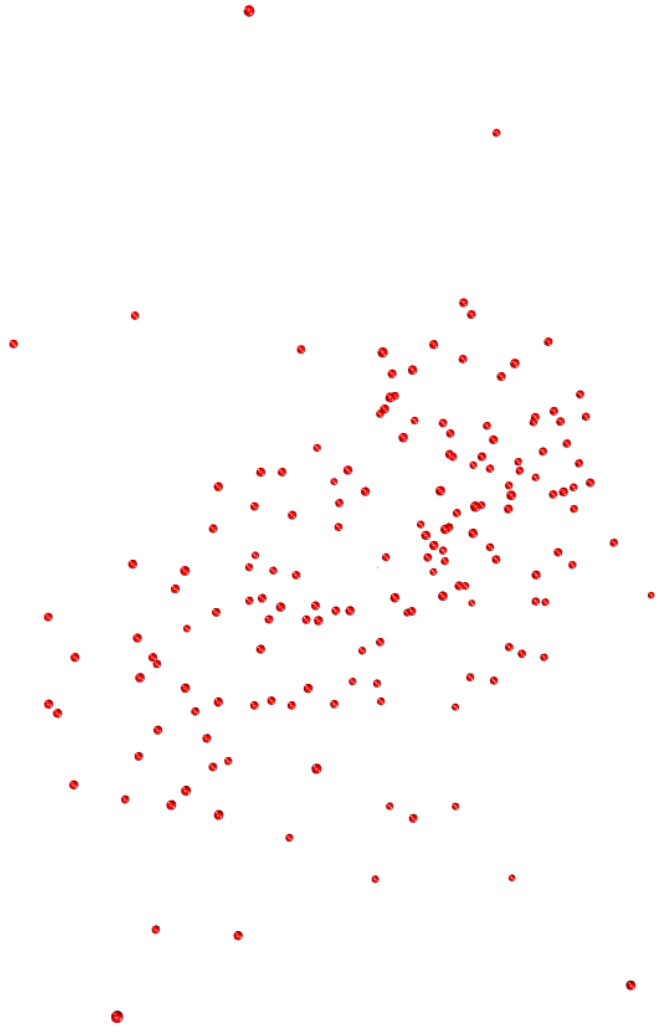
Multiple-output regression quantiles (illustration)



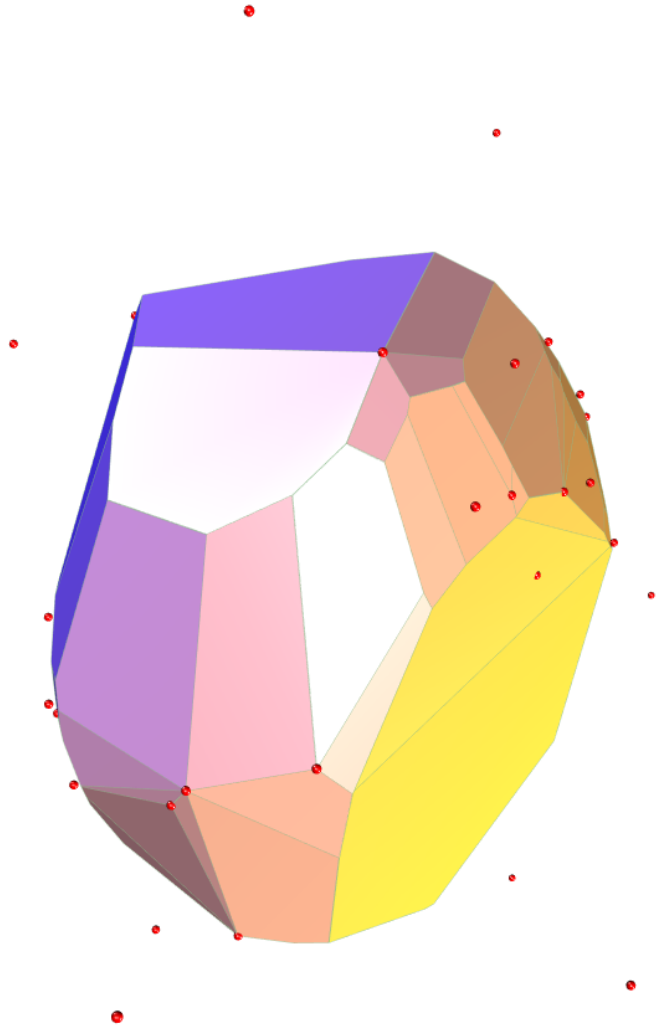
Multiple-output regression quantiles (illustration)



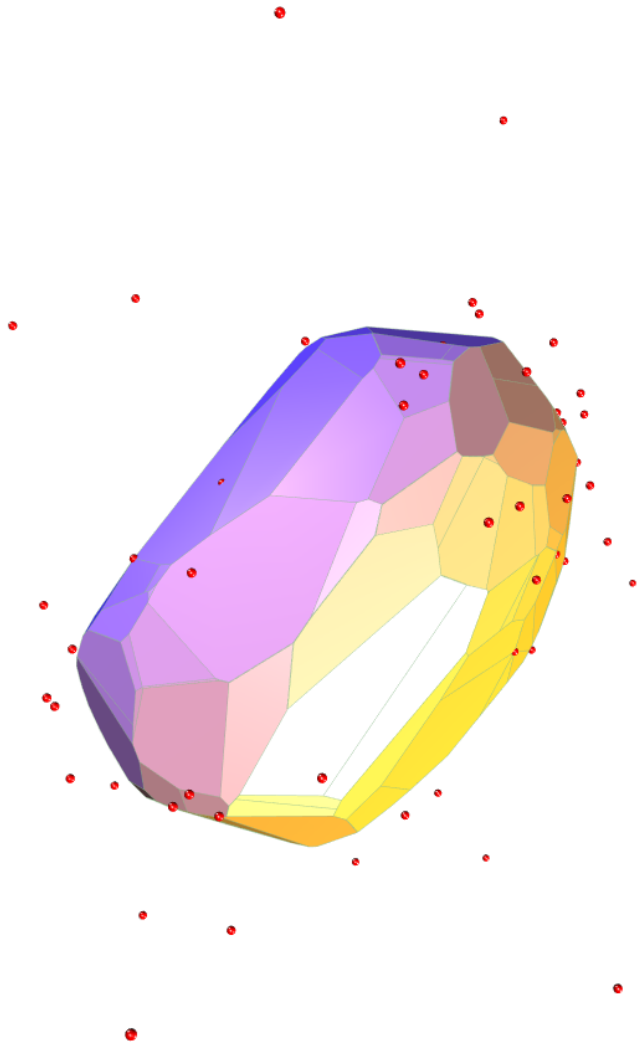
Tukey (=halfspace, location) depth region



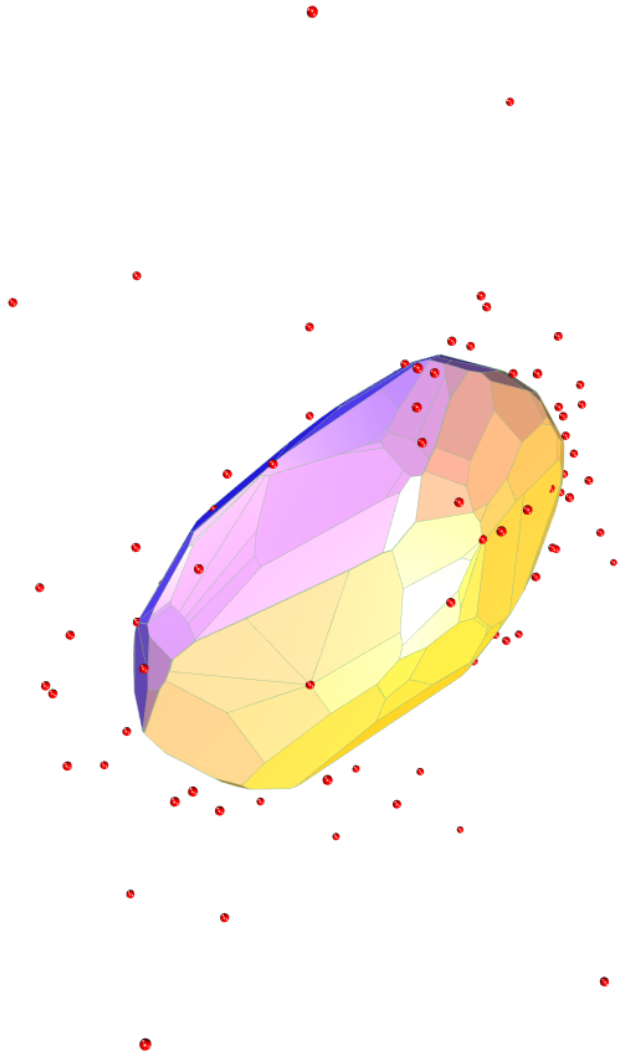
Tukey (=halfspace, location) depth region: $\alpha = 2/161$



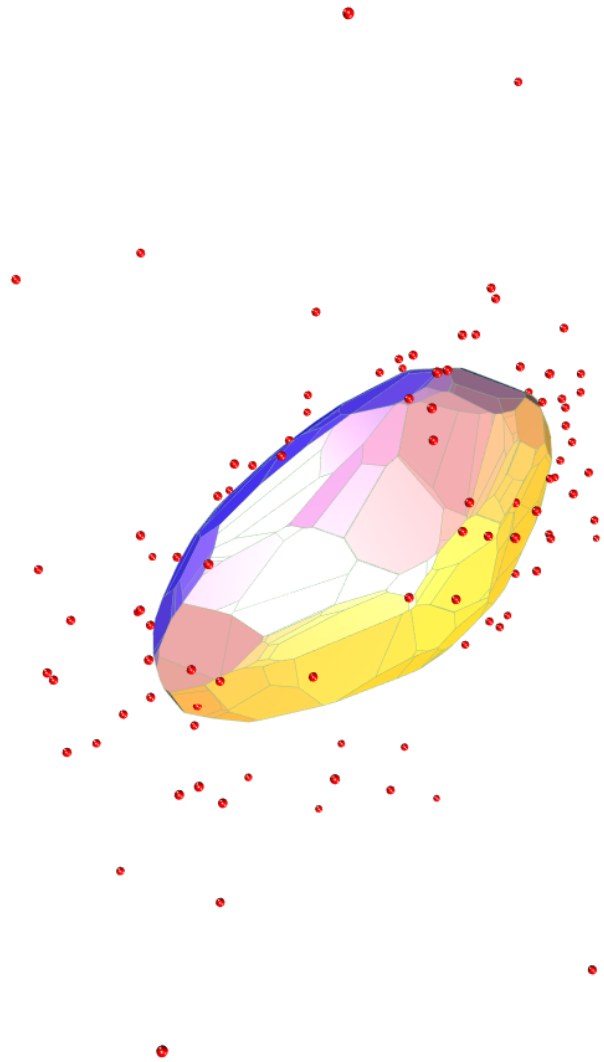
Tukey (=halfspace, location) depth region: $\alpha = 5/161$



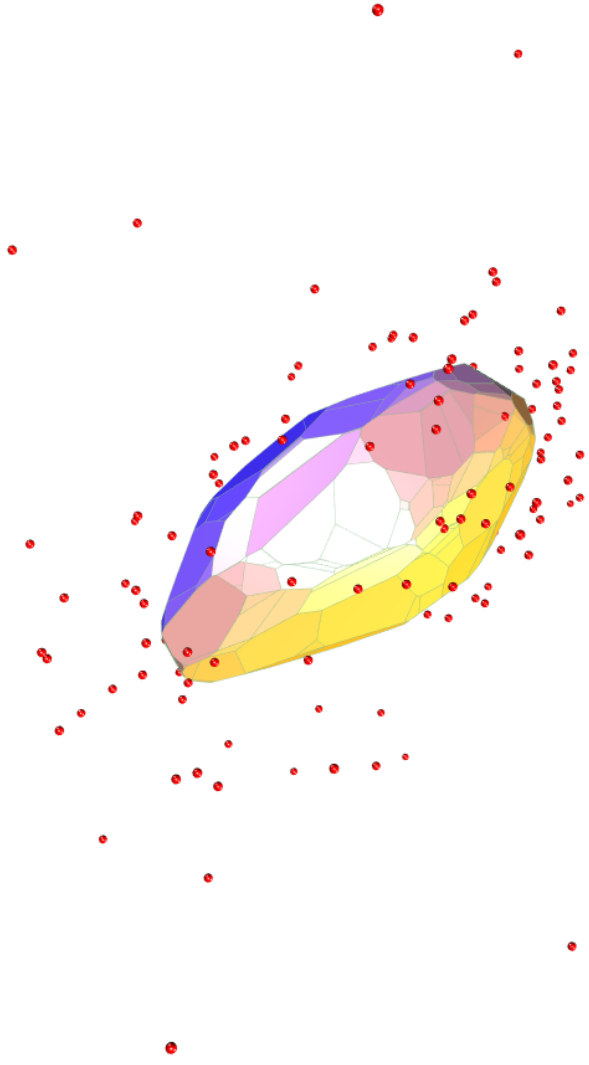
Tukey (=halfspace, location) depth region: $\alpha = 9/161$



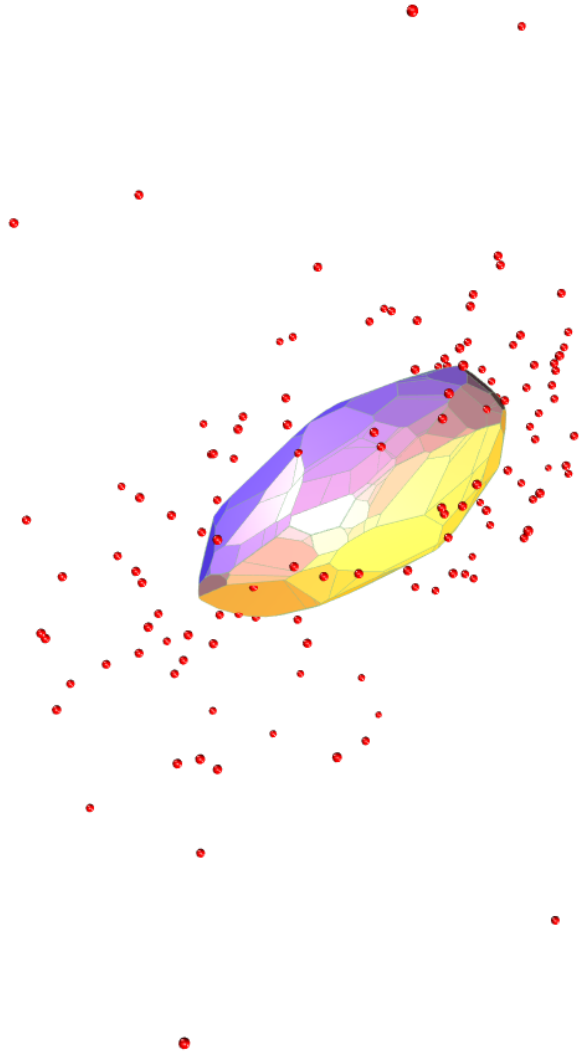
Tukey (=halfspace, location) depth region: α =13/161



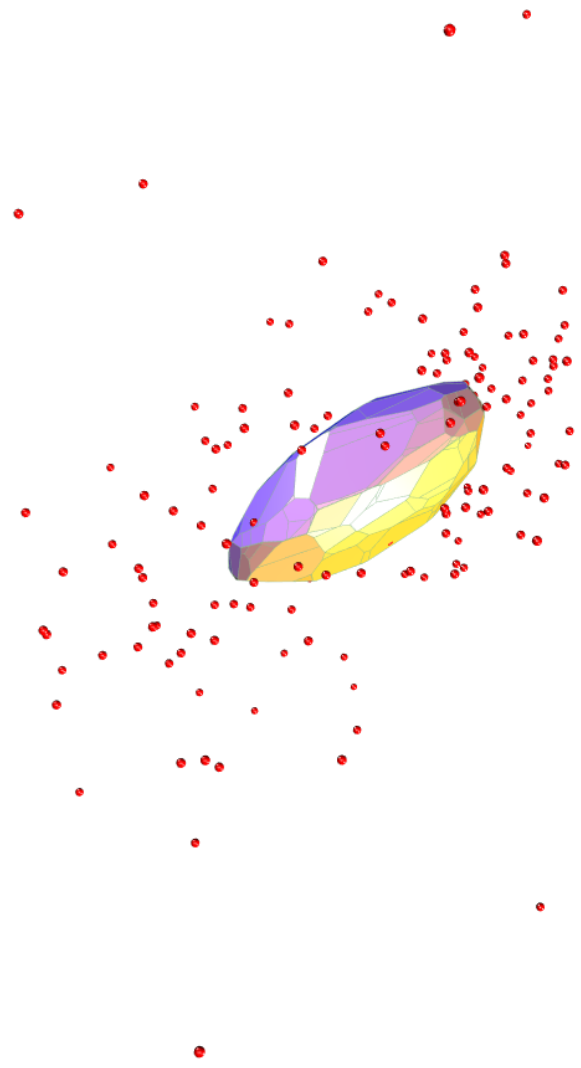
Tukey (=halfspace, location) depth region: α =17/161



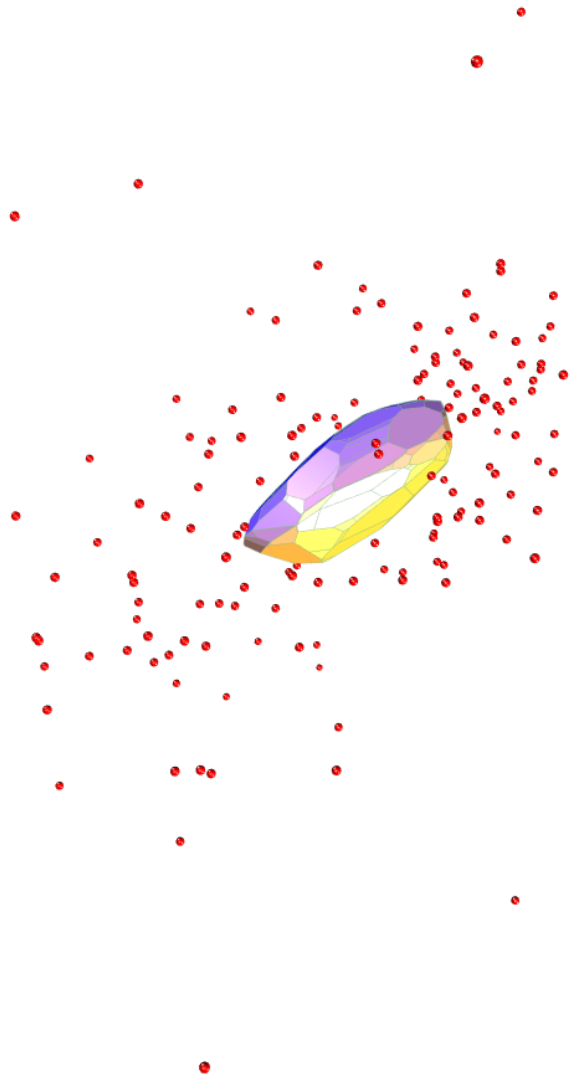
Tukey (=halfspace, location) depth region: α =25/161



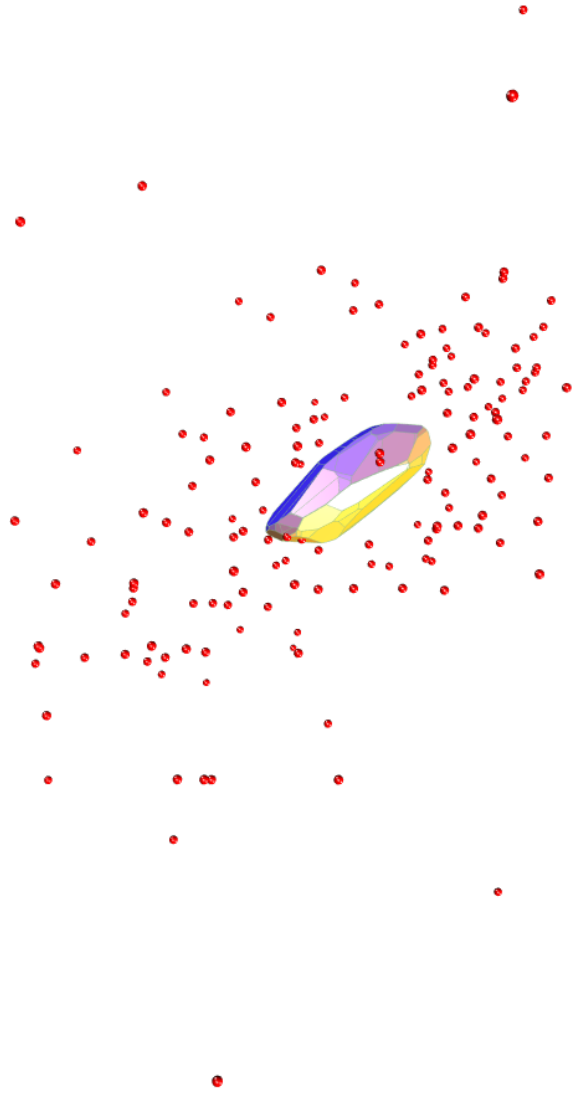
Tukey (=halfspace, location) depth region: α =33/161



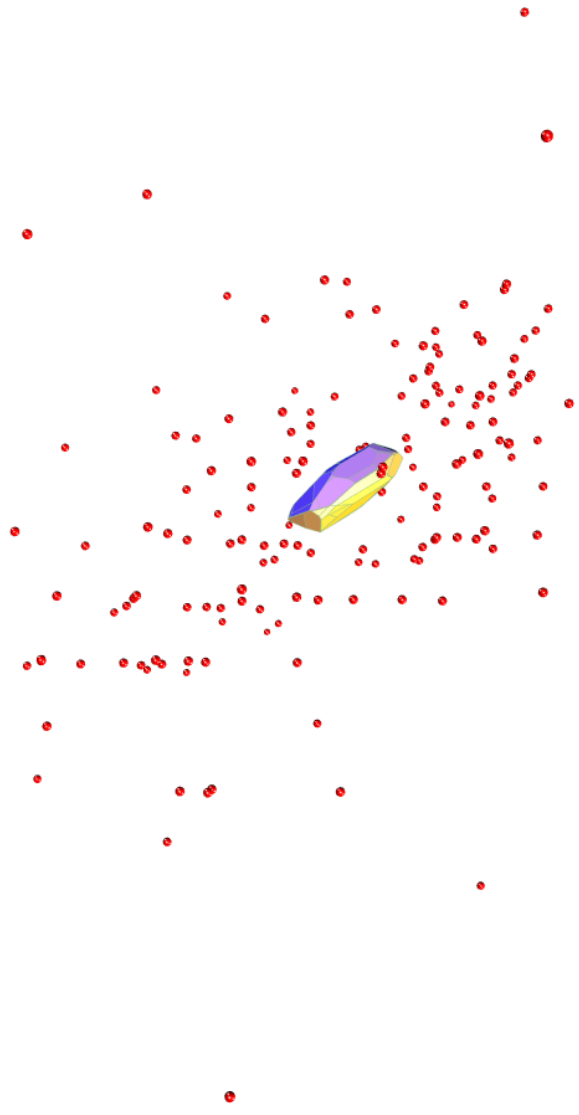
Tukey (=halfspace, location) depth region: α =41/161



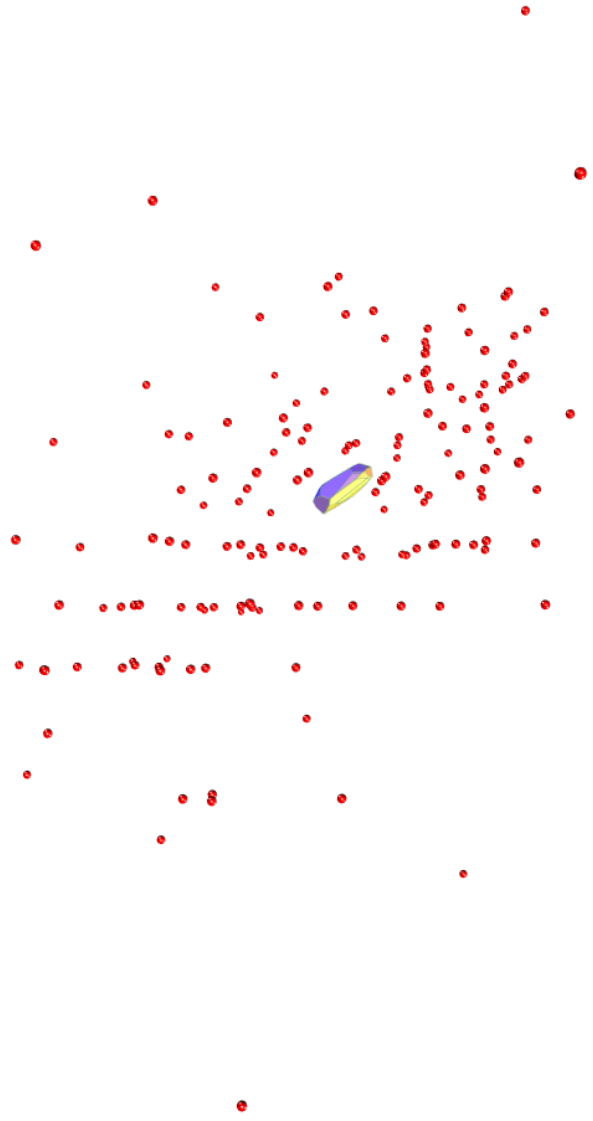
Tukey (=halfspace, location) depth region: α =49/161



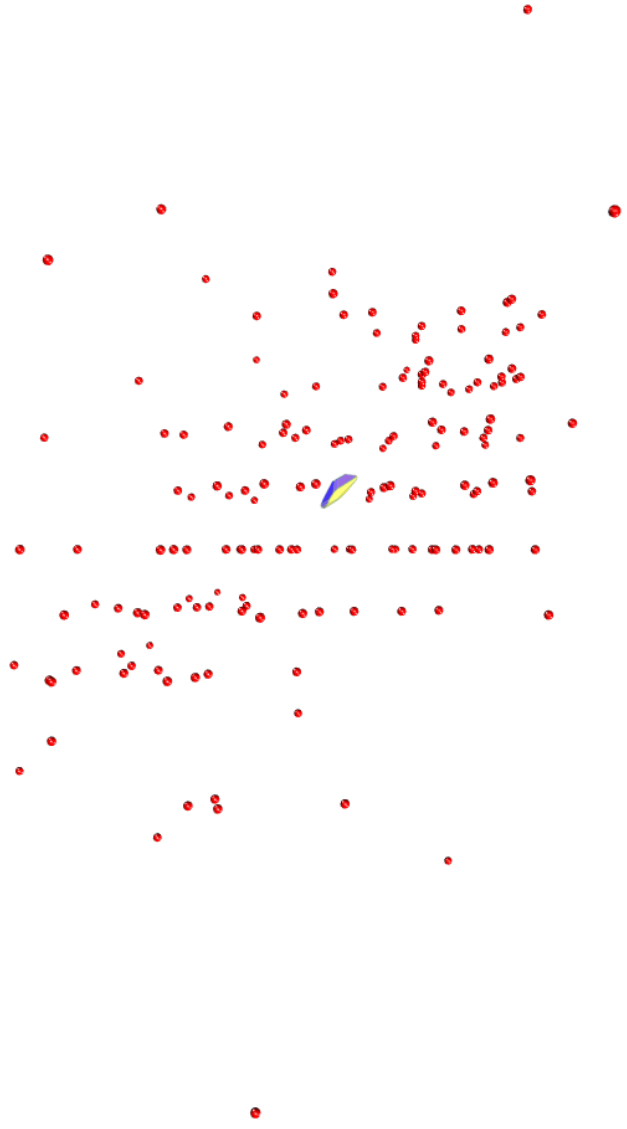
Tukey (=halfspace, location) depth region: α =57/161



Tukey (=halfspace, location) depth region: α =65/161



Tukey (=halfspace, location) depth region: α =68/161



Contents

The concept of data depth

Some examples
- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth
- Properties
- Trimmed regions

Classes of depth functions
- Depths satisfying the projection property

A concept of depth for the functional framework
- Functional outlier detection

References

Contents

The concept of data depth

Some examples

- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth

- Properties
- Trimmed regions

Classes of depth functions

- Depths satisfying the projection property

A concept of depth for the functional framework

- Functional outlier detection

References

Projection property (Dyckerhoff, 2004)

- Let D be a depth. D satisfies the (weak) projection property, if for each point $\boldsymbol{x} \in \mathbb{R}^d$ and each random vector X it holds:

$$D(\boldsymbol{x}|X) = \inf_{\boldsymbol{r} \in S^{d-1}} D(\boldsymbol{r}^T \boldsymbol{x}|\boldsymbol{r}^T X)\,.$$

- If a depth satisfies the weak projection property, then for each direction $\boldsymbol{r} \in S^{d-1}$ we have:

$$D(\boldsymbol{x}|X) \leq D(\boldsymbol{r}^T \boldsymbol{x}|\boldsymbol{r}^T X)\,.$$

- Let D be a depth which satisfies the weak projection property. If in addition for every $z \in \mathbb{R}$ and $\boldsymbol{r} \in S^{d-1}$ there exists a $\boldsymbol{x} \in \mathbb{R}^d$ such that $\boldsymbol{r}^T \boldsymbol{x} = z$ and

$$D(\boldsymbol{x}|X) = D(\boldsymbol{r}^T \boldsymbol{x}|\boldsymbol{r}^T X)\,,$$

  then we say that D satisfies the strong projection property.

Projection property (Dyckerhoff, 2004)

- Let D be a depth, then
  1. D satisfies the weak projection property if and only if for every random vector X and $\alpha \geq 0$ holds

$$D_\alpha(X) = \bigcap_{\boldsymbol{r} \in S^{d-1}} \boldsymbol{r}^{-1}\big(D_\alpha(\boldsymbol{r}^T X)\big)\,;$$

  2. D satisfies the strong projection property if and only if for every random vector X, $\alpha \geq 0$, and $\boldsymbol{r} \in S^{d-1}$ holds the above equation and in addition

$$\boldsymbol{r}^T D_\alpha(X) = D_\alpha(\boldsymbol{r}^T X)\,.$$

- If a depth D is quasiconcave (*i.e.* satisfies D4con), we say that D is a convex depth.
- Let D be a depth that satisfies the weak projection property. Then D is a convex depth.

Projection property (Dyckerhoff, 2004)

- For a compact convex subset $K \in \mathbb{R}^d$ the support function $h_K : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$h_K(\boldsymbol{r}) = \max\{\boldsymbol{r}^T\boldsymbol{x} : \boldsymbol{x} \in K\}\,.$$

- Let D be a depth. D satisfies the strong projection property if and only if for every random vector X and direction $\boldsymbol{r} \in S^{d-1}$

$$h_{D_\alpha(X)}(\boldsymbol{r}) = h_{D_\alpha(\boldsymbol{r}^T X)}(1)\,.$$

- Let D^1 be a univariate depth. If D is defined by

$$D(\boldsymbol{x}|X) = \inf_{\boldsymbol{r}\in S^{d-1}} D^1(\boldsymbol{r}^T\boldsymbol{x}|\boldsymbol{r}^T X)\,,$$

  then D is a multivariate convex depth that satisfies the weak projection property.
- Satisfaction of the property by depths:
  - Mahalanobis depth satisfies the strong projection property.
  - Halfspace depth satisfies the weak projection property.
  - Zonoid depth satisfies the strong projection property.
  - Projetion depth satisfies the weak projection property.

Contents

The concept of data depth

Some examples
- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

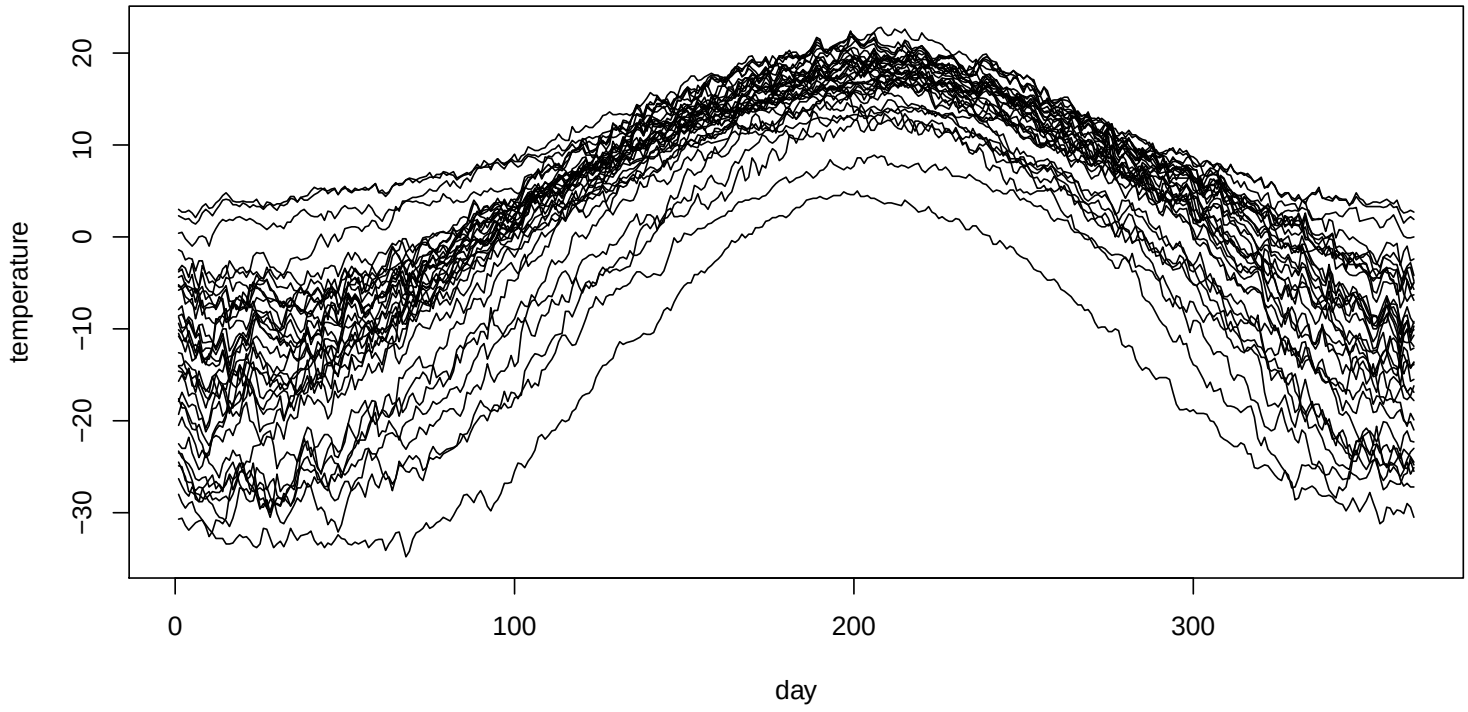
Halfspace depth
- Properties
- Trimmed regions

Classes of depth functions
- Depths satisfying the projection property

A concept of depth for the functional framework
- Functional outlier detection

References

The functional data depth: an example

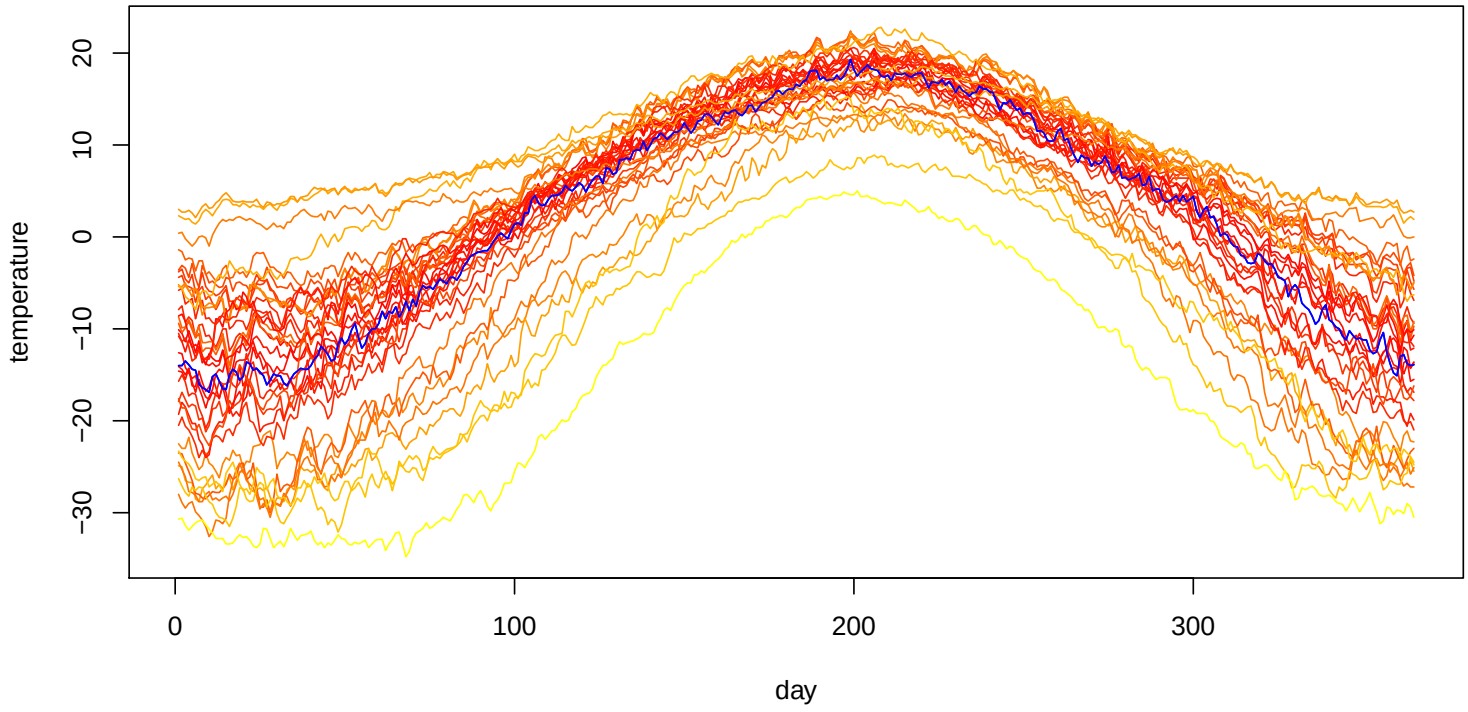


Let X be a stochastic process with continuous paths defined on $[0, 1]$, and $\boldsymbol{x}$ its realization. Then:

$$D(\boldsymbol{x}|X) = \int_0^1 D\big(\boldsymbol{x}(v)|X(v)\big)\,dv.$$

see Fraiman, Muniz (2001); also López-Pintado, Romo (2011).

The functional data depth: an example



Let X be a stochastic process with continuous paths defined on $[0, 1]$, and $\boldsymbol{x}$ its realization. Then:

$$D(\boldsymbol{x}|X) = \int_0^1 \min\{F_{X(v)}\big(\boldsymbol{x}(v)\big), 1 - F_{X(v)}\big(\boldsymbol{x}(v)^-\big)\}dv.$$

see Fraiman, Muniz (2001); also López-Pintado, Romo (2011).

Contents

The concept of data depth

Some examples

- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth

- Properties
- Trimmed regions

Classes of depth functions

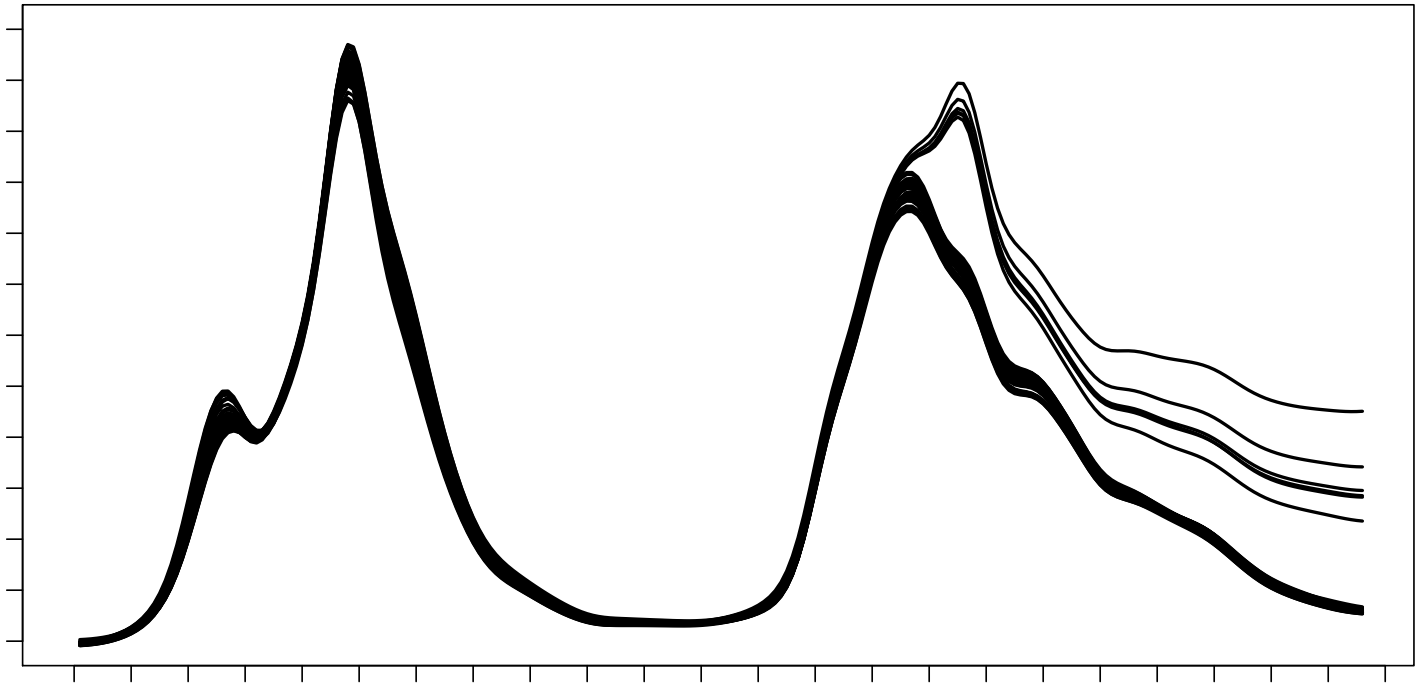
- Depths satisfying the projection property

A concept of depth for the functional framework

- Functional outlier detection

References

Detection of functional outliers

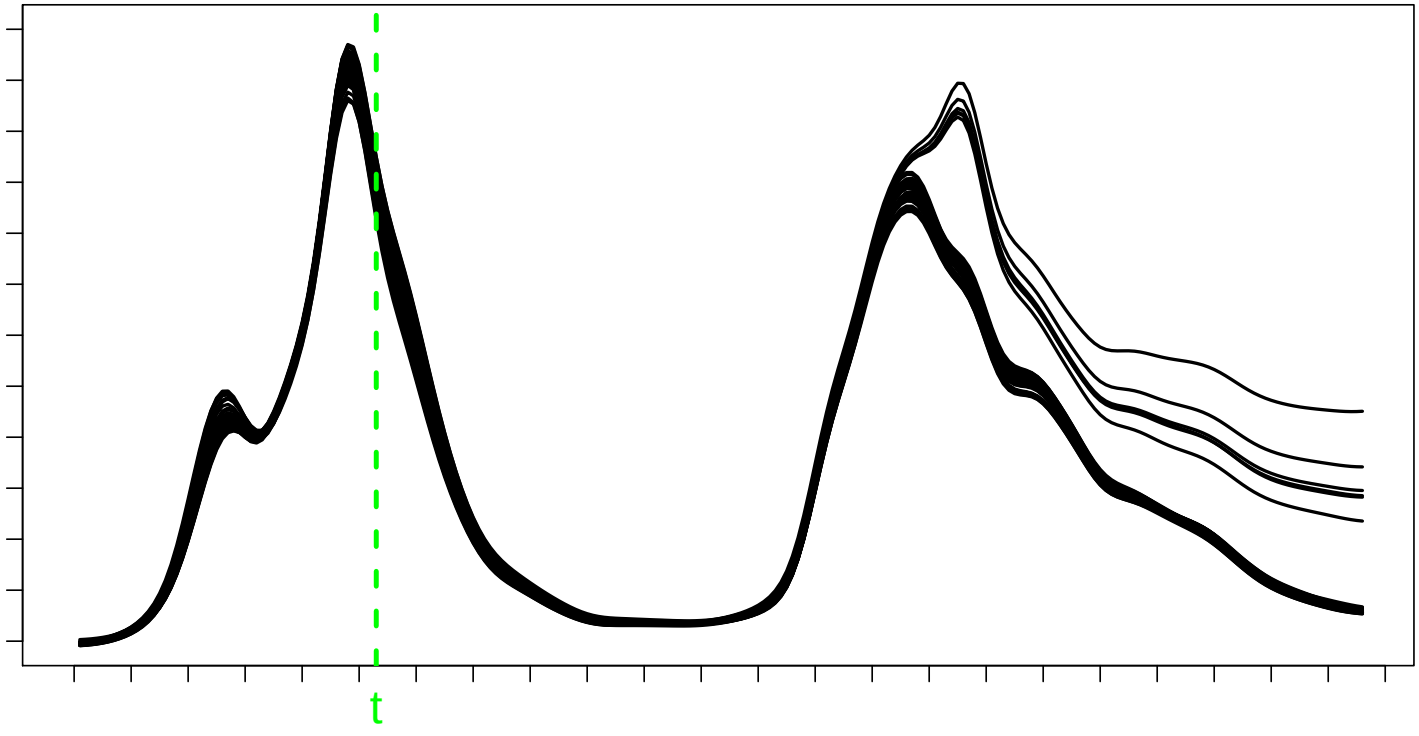


- Functional (projection) depth of $\boldsymbol{f}$ w.r.t. $\mathcal{F} = \{\boldsymbol{f}_i\}_{i=1}^n$:

$$D^{prj}(\boldsymbol{f}|\mathcal{F}) = \int_{t_{\min}}^{t^{\max}} D^{prj}\left(\boldsymbol{f}(t)|\mathcal{F}(t)\right) dt\,,$$

where $D_{prj}(\cdot|\cdot)$ is the projection depth, as above.

Detection of functional outliers

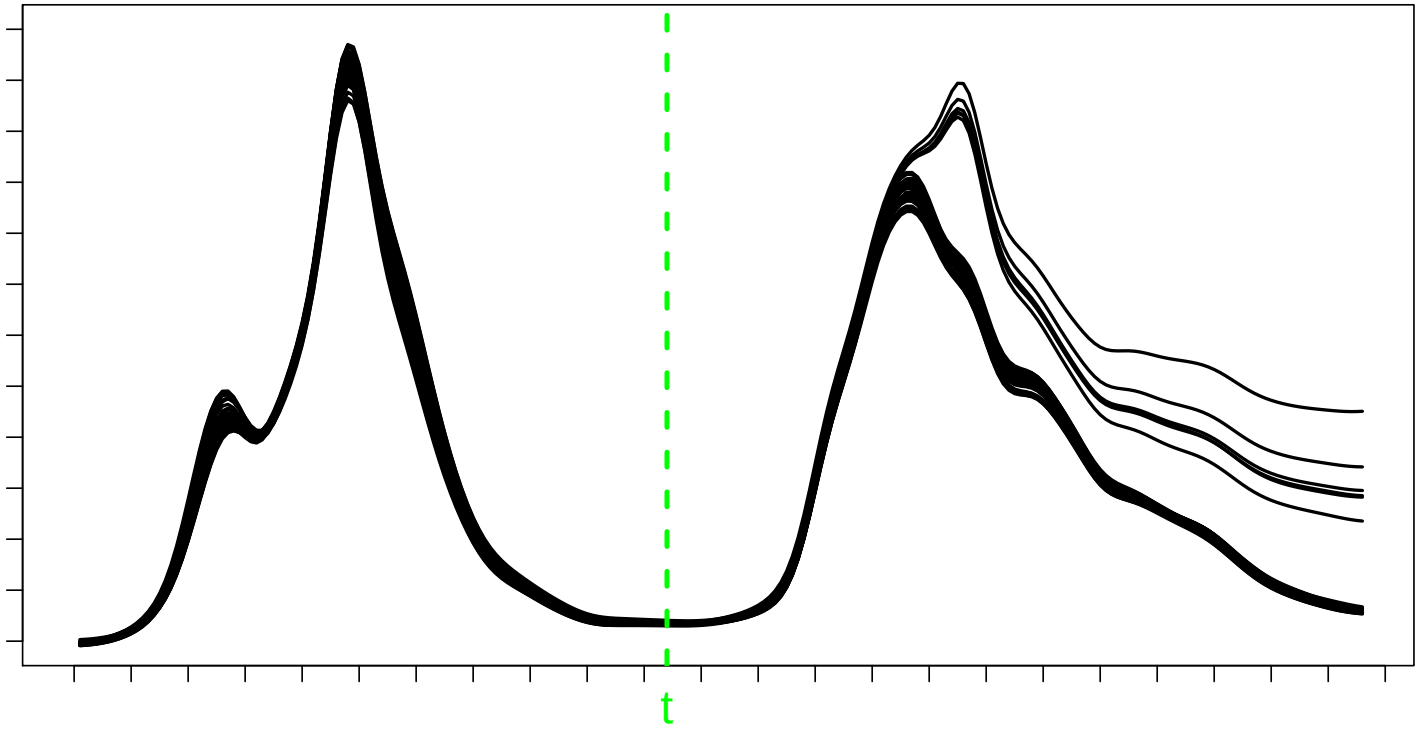


- Functional (projection) depth of $\boldsymbol{f}$ w.r.t. $\mathcal{F} = \{\boldsymbol{f}_i\}_{i=1}^n$:

$$D^{prj}(\boldsymbol{f}|\mathcal{F}) = \int_{t_{\min}}^{t^{\max}} D^{prj}\left(\boldsymbol{f}(t)|\{\boldsymbol{f}_1(t), ..., \boldsymbol{f}_n(t)\}\right) dt\,,$$

where $D_{prj}(\cdot|\cdot)$ is the projection depth, as above.

Detection of functional outliers

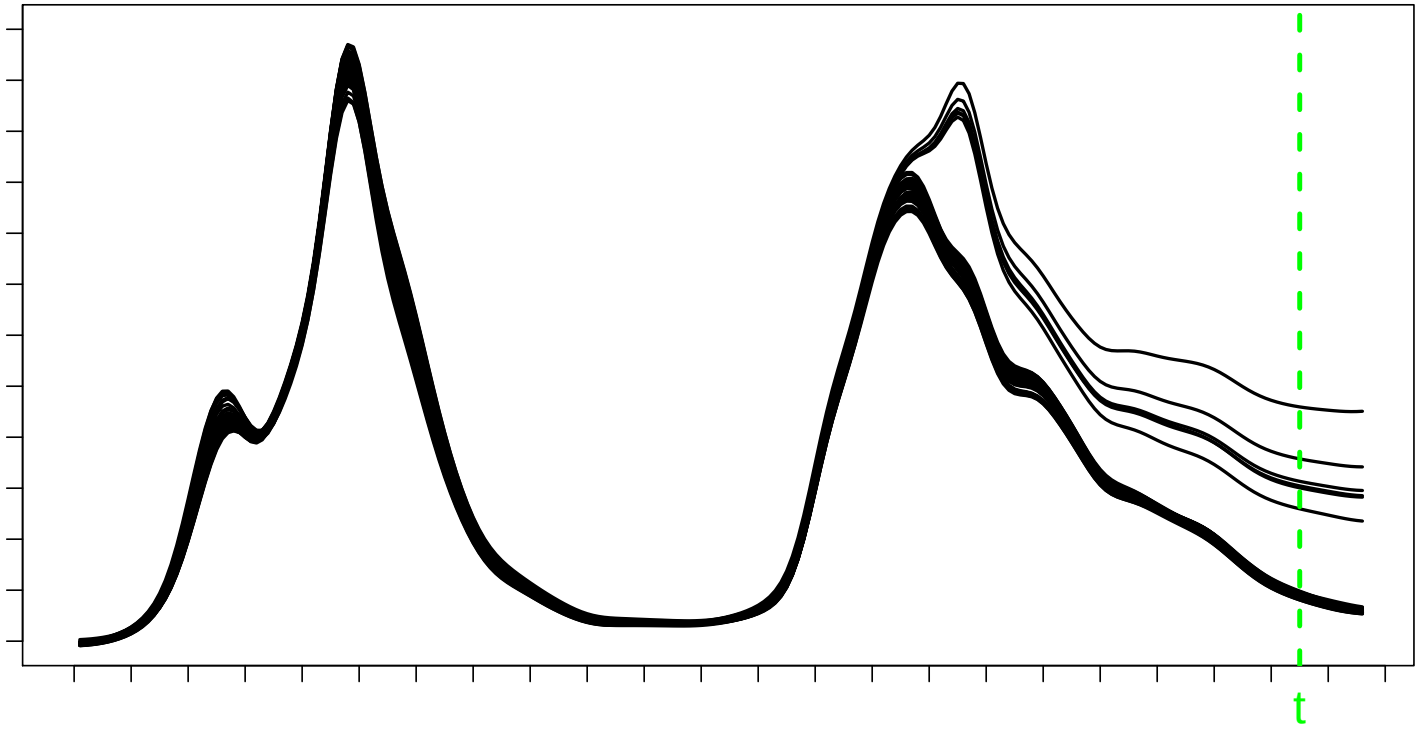


- Functional (projection) depth of $\boldsymbol{f}$ w.r.t. $\mathcal{F} = \{\boldsymbol{f}_i\}_{i=1}^n$:

$$D^{prj}(\boldsymbol{f}|\mathcal{F}) = \int_{t_{\min}}^{t^{\max}} D^{prj}\left(\boldsymbol{f}(t)|\{\boldsymbol{f}_1(t), ..., \boldsymbol{f}_n(t)\}\right) dt\,,$$

where $D_{prj}(\cdot|\cdot)$ is the projection depth, as above.

Detection of functional outliers

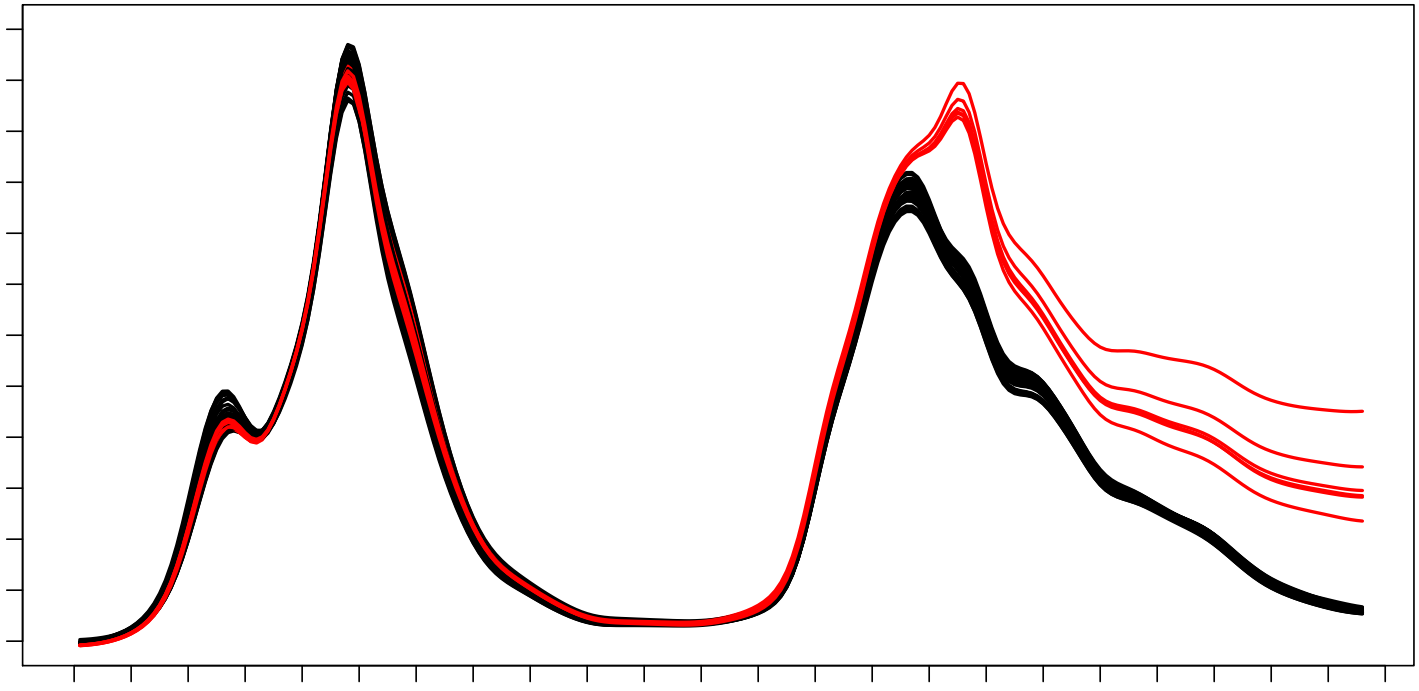


- Functional (projection) depth of $\boldsymbol{f}$ w.r.t. $\mathcal{F} = \{\boldsymbol{f}_i\}_{i=1}^n$:

$$D^{prj}(\boldsymbol{f}|\mathcal{F}) = \int_{t_{\min}}^{t^{\max}} D^{prj}\left(\boldsymbol{f}(t)|\{\boldsymbol{f}_1(t), ..., \boldsymbol{f}_n(t)\}\right) dt\,,$$

where $D_{prj}(\cdot|\cdot)$ is the projection depth, as above.

Detection of functional outliers



- Functional (projection) depth of $\boldsymbol{f}$ w.r.t. $\mathcal{F} = \{\boldsymbol{f}_i\}_{i=1}^n$:

$$D^{prj}(\boldsymbol{f}|\mathcal{F}) = \int_{t_{\min}}^{t^{\max}} D^{prj}\left(\boldsymbol{f}(t)|\{\boldsymbol{f}_1(t), ..., \boldsymbol{f}_n(t)\}\right) dt\,,$$

  where $D_{prj}(\cdot|\cdot)$ is the projection depth, as above.
- Label $\boldsymbol{f}$ as outlier if $D^{prj}(\boldsymbol{f}; \mathcal{F}) < \min(D^{prj})$.

Contents

The concept of data depth

Some examples
- Mahalanobis depth
- Spatial depth
- Projection depth
- Zonoid depth

Depth-based outlier detection

Halfspace depth
- Properties
- Trimmed regions

Classes of depth functions
- Depths satisfying the projection property

A concept of depth for the functional framework
- Functional outlier detection

References

Thank you for your attention! Questions?

- Brunel VE (2019). Concentration of the Empirical Level Sets of Tukey's Halfspace Depth. *Probability Theory and Related Fields*, **173**, 1165–1196.
- Burr MA, Fabrizio RJ (2017). Uniform Convergence Rates for Halfspace Depth. *Statistics and Probability Letters*, **124**, 33–40.
- Chaudhuri P (1996). On a Geometric Notion of Quantiles for Multivariate Data. *Journal of the American Statistical Association*, **91**, 862–872.
- Donoho D (1982). *Breakdown Properties of Multivariate Location Estimators.* Ph.D. thesis, Harvard University.
- Donoho DL, Gasko M (1992). Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. *The Annals of Statistics*, **20**, 1803–1827.
- Dyckerhoff R (2004). Data Depths Satisfying the Projection Property. *AStA – Advances in Statistical Analysis*, **88**, 163–190.
- Dyckerhoff R (2017). Convergence of Depths and Depth-Trimmed Regions. *arXiv:1611.08721*.
- Dyckerhoff R, Mosler K (2011). Weighted-Mean Trimming of Multivariate Data. *Journal of Multivariate Analysis*, **102**, 405–421.
- Dyckerhoff R, Mosler K, Koshevoy G (1996). Zonoid Data Depth: Theory and Computation. In A Prat (ed.), *COMPSTAT '96 – Proceedings in Computational Statistics*, pp. 235–240. Springer.

Thank you for your attention! Questions?

- Dyckerhoff R, Mozharovskyi P (2016). Exact Computation of the Halfspace Depth. *Computational Statistics and Data Analysis*, **98**, 19–30.
- Hallin M, Paindaveine D, Siman M (2010). Multivariate Quantiles and Multiple-Output Regression Quantiles. From L1 Optimization to Halfspace Depth. *The Annals of Statistics*, **38**, 635–669.
- Koltchinskii V (1997). M-Estimation, Convexity and Quantiles. *The Annals of Statistics*, **25**, 435–477.
- Kong L, Mizera I (2012). Quantile Tomography: Using Quantiles with Multivariate Data. *Statistica Sinica*, **22**, 1589–1610.
- Koshevoy G (2002). The Tukey Depth Characterizes the Atomic Measure. *Journal of Multivariate Analysis*, **83**, 360–364.
- Koshevoy G (2003). Lift-Zonoid and Multivariate Depths. In R Dutter, P Filzmoser, U Gather, P Rousseeuw (eds.), *Developments in Robust Statistics*, pp. 194–202. Physica-Verlag, Heidelberg.
- Koshevoy G, Mosler K (1997). Zonoid Trimming for Multivariate Distributions. *The Annals of Statistics*, **25**, 1998–2017.
- Lange T, Mosler K, Mozharovskyi P (2014). Fast nonparametric classification based on data depth. *Statistical Papers*, **55**, 49–69.
- Li J, Cuesta-Albertos JA, Liu RY (2012). *DD*-Classifier: Nonparametric Classification Procedure Based on *DD*-Plot. *Journal of the American Statistical Association*, **107**, 737–753.

Thank you for your attention! Questions?

- Liu RY (1990). On a Notion of Data Depth Based on Random Simplices. *The Annals of Statistics*, **18**, 405–414.
- Mahalanobis PC (1936). On the Generalized Distance in Statistics. *Proceedings of the National Institute of Sciences of India*, **12**, 49–55.
- Masse JC (2002). Asymptotics for the Tukey Median. *Journal of Multivariate Analysis*, **81**, 286–300.
- Masse JC (2004). Asymptotics for the Tukey Depth Process, with an Application to a Multivariate Trimmed Mean. *Bernoulli*, **10**, 397–419.
- Mizera I (2002). On Depth and Deep Points: A Calculus. *The Annals of Statistics*, **30**, 1681–1736.
- Mosler K (2002). *Multivariate Dispersion, Central Regions, and Depth: The Lift Zonoid Approach.* Springer-Verlag, New York.
- Mosler K (2013). Depth Statistics. In C Becker, R Fried, S Kuhnt (eds.), *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*, pp. 17–34. Springer. Berlin.

  Mozharovskyi P, Josse J, Husson F (2018). Nonparametric imputation by data depth. *Journal of the American Statistical Association*, in press.
- Nieto-Reyes A, Battey H (2016). A Topologically Valid Definition of Depth for Functional Data. *Statistical Science*, **31**, 61–79.

Thank you for your attention! Questions?

- Paidaveine D, Siman M (2012). Computing Multiple-Output Regression Quantile Regions. *Computational Statistics and Data Analysis*, **56**, 840–853.
- Paidaveine D, Siman M (2012). Computing Multiple-Output Regression Quantile Regions from Projection Quantiles. *Computational Statistics*, **27**, 29–49.
- Rousseeuw PJ, Ruts I (1999). The Depth Function of a Population Distribution. *Metrika*, **49**, 213–244.
- Stahel W (1981). *Robust Estimation: Infinitesimal Optimality and Covariance Matrix Estimators (In German).* Ph.D. thesis, Swiss Federal Institute of Technology in Zurich.
- Struyf AJ, Rousseeuw PJ (1999). Halfspace Depth and Regression Depth Characterize the Empirical Distribution. *Journal of Multivariate Analysis*, **69**, 135–153.
- Tukey JW (1975). Mathematics and the Picturing of Data. In R James (ed.), *Proceedings of the International Congress of Mathematicians*, volume 2, pp. 523–531. Canadian Mathematical Congress.
- Vardi Y, Zhang C (2000). The Multivariate L_1-Median and Associated Data Depth. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 1423–1426.
- Zuo Y, Serfling R (2000). General Notions of Statistical Depth Function. *The Annals of Statistics*, **28**, 461–482.