Gov 2000 - 10. Troubleshooting the Linear Model

Matthew Blackwell

November 24, 2015

1. Nonnormality of the errors

2. Nonlinearity of the regression function

3. Outliers, leverage points, and influential observations

Where are we? Where are we going?

- Last few weeks: estimation and inference for the linear model under Gauss-Markov assumptions (and sometimes conditional Normality)
- This week: what happens when the assumptions fail? Can we tell? Can we fix it?
- Next weeks: more of the same

Review of the OLS assumptions

1. Linearity: $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$
2. Random/iid sample: $(y_i, \mathbf{x}_i')$ are a iid sample from the population.
3. No perfect collinearity: $\mathbf{X}$ is an $n \times (k+1)$ matrix with rank $k+1$
4. Zero conditional mean: $\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbf{0}$
5. Homoskedasticity: $\text{var}(\mathbf{u}|\mathbf{X}) = \sigma_u^2 \mathbf{I}_n$
6. Normality: $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_n)$

- 1-4 give us unbiasedness/consistency
- 1-5 are the Gauss-Markov, allow for large-sample inference
- 1-6 allow for small-sample inference

Violations of the assumptions

1. Nonlinearity
   - Result: biased/inconsistent estimates
   - Diagnose: scatterplots, added variable plots, component-plus-residual plots
   - Correct: transformations, polynomials, splines
2. iid/random sample
   - Result: no bias with appropriate alternative assumptions (structured dependence)
   - Result (ii): violations imply heteroskedasticity
   - Result (iii): outliers from different distributions can cause inefficiency/bias
   - Diagnose/Correct: next week!
3. Perfect collinearity
   - Result: can't run OLS
   - Diagnose/correct: drop one collinear term

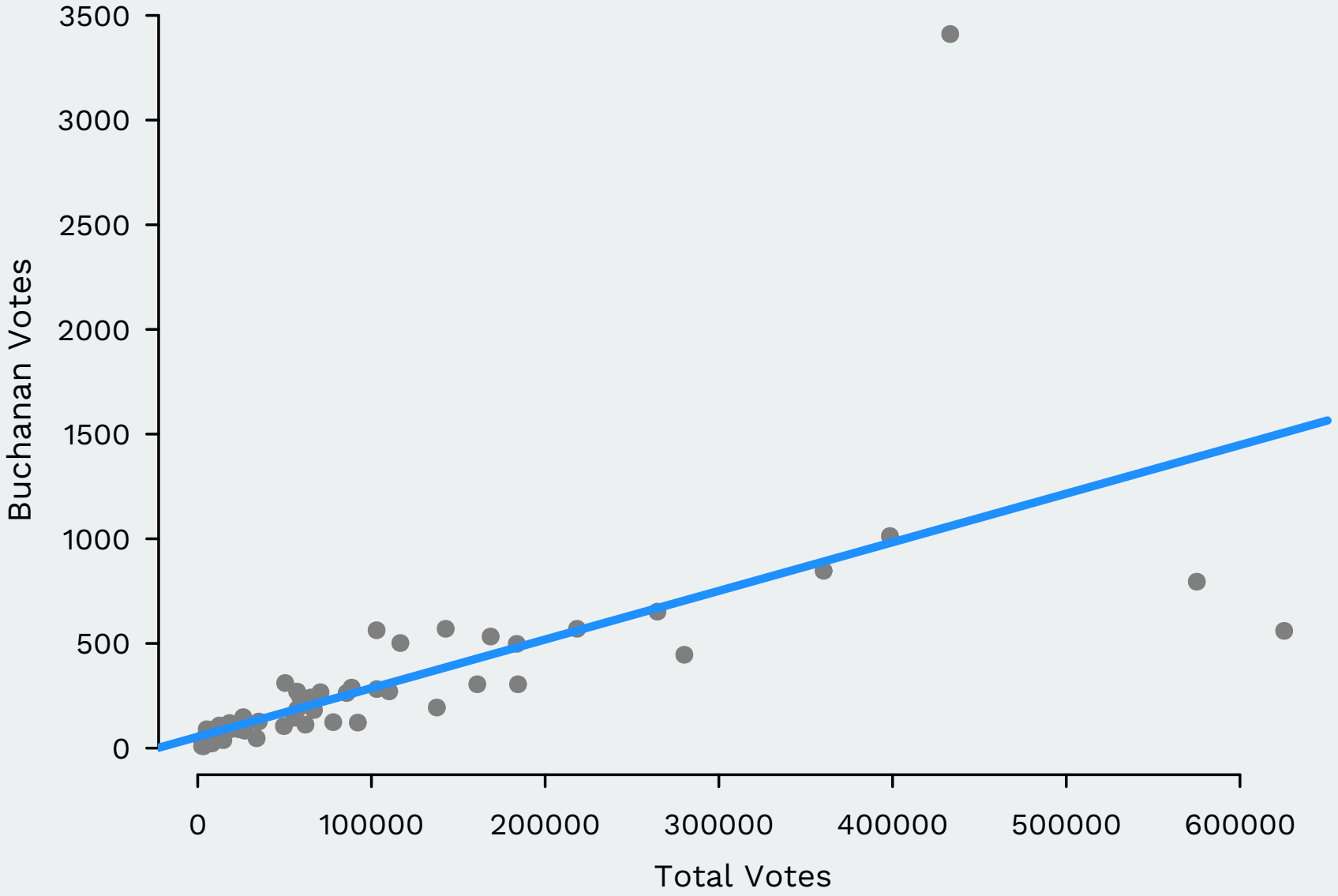
Violations of the assumptions (ii)

4. Zero conditional mean error
   - Result: biased/inconsistent estimates
   - Diagnose: very difficult
   - Correct: instrumental variables, fixed effects, regression discontinuity (Gov 2002)
5. Heteroskedasticity
   - Result: SEs are biased (usually downward)
   - Diagnose/correct: next week!
6. Non-Normality
   - Result: critical values for t and F tests wrong
   - Diagnose: checking the (studentized) residuals, QQ-plots, etc
   - Correct: transformations, add variables to $\mathbf{X}$, different model

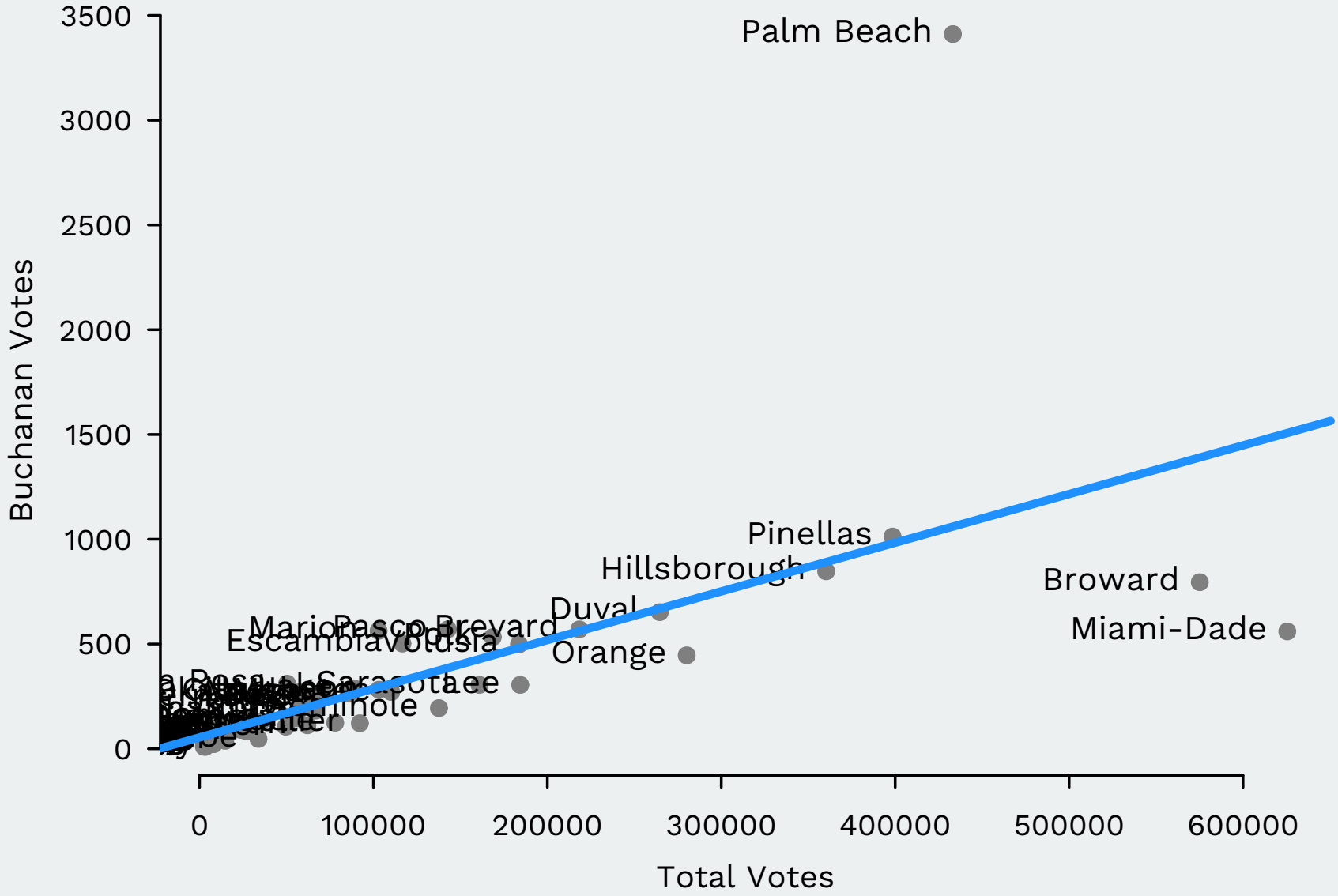
Example: Buchanan votes in Florida, 2000

- 2000 Presidential election in FL (Wand et al., 2001, APSR)

Example: Buchanan votes in Florida, 2000



Example: Buchanan votes in Florida, 2000



1/ Nonnormality of the errors

Review of the Normality assumption

- In matrix notation:

$$\mathbf{u}|\mathbf{X} \sim \mathcal{N}(0, \sigma_u^2 \mathbf{I})$$

- Equivalent to:

$$u_i|\mathbf{x}_i' \sim \mathcal{N}(0, \sigma_u^2)$$

- Fix $\mathbf{x}_i'$ and the distribution of errors should be Normal

Consequences of non-Normal errors?

- In small samples:
  - Sampling distribution of $\widehat{\boldsymbol{\beta}}$ will not be Normal
  - Test statistics will not have t or F distributions
  - Probability of Type I error will not be α
  - $1 - \alpha$ confidence interval will not have $1 - \alpha$ coverage
- In large samples:
  - Sampling distribution of $\widehat{\boldsymbol{\beta}} \approx$ Normal by the CLT
  - Test statistics will be $\approx t$ or F by the CLT
  - Probability of Type I error $\approx \alpha$
  - $1 - \alpha$ confidence interval will have $\approx 1 - \alpha$ coverage
- The n needed for approximation to hold depends on how non-Normal the data are

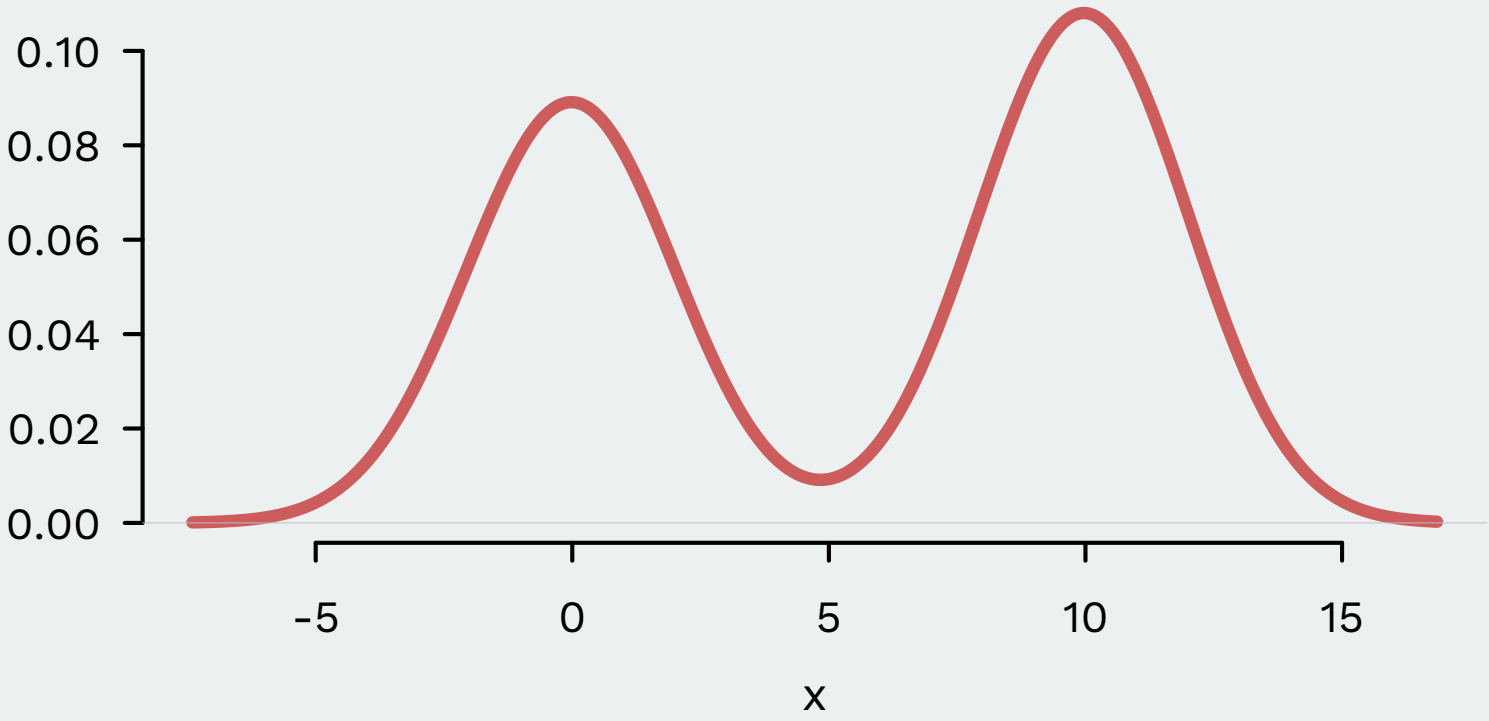
Marginal versus conditional

- Be careful with this assumption: distribution of the error, not the distribution of y_i
- The marginal distribution of y_i can be non-Normal even if the conditional distribution is Normal!

Is this a violation?

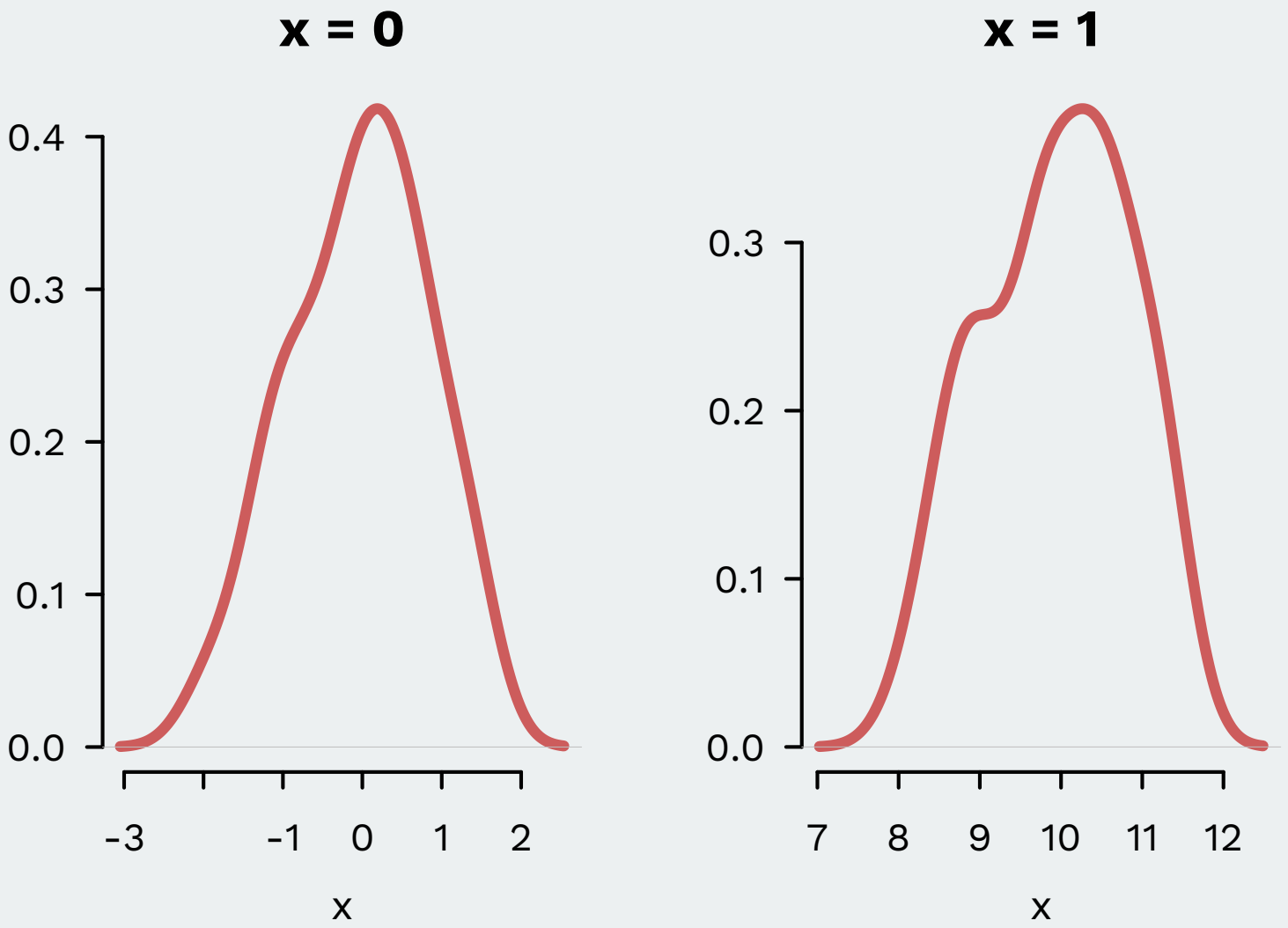
- For example, this looks bad:

```
x <- rbinom(100, 1, 0.5)
y <- 10 * x + rnorm(100, 0, 1)
plot(density(y), lwd = 3, col = "indianred", las = 1, xlab = "x", main = ""
    bty = "n", ylab = "")
```



Is this a violation?

- But if we look at the conditional distributions, things look better:



How to diagnose?

- Assumption is about unobserved $\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$
- We can only observe residuals, $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$
- If distribution of residuals $\approx$ distribution of errors, we could check residuals
- But this is actually not true—the distribution of the residuals is complicated

Hat matrix

- First we need to define an important matrix $\mathbf{H} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \\ &= \mathbf{y} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \\ &\equiv \mathbf{y} - \mathbf{H}\mathbf{y} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{y}\end{aligned}$$

- $\mathbf{H}$ is the hat matrix because it puts the "hat" on $\mathbf{y}$:

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$$

  - $\mathbf{H}$ is an $n \times n$ symmetric matrix
  - $\mathbf{H}$ is idempotent: $\mathbf{H}\mathbf{H} = \mathbf{H}$

Relating the residuals to the errors

- Possible to show that residuals $\hat{\mathbf{u}}$ are a linear function of the errors, $\mathbf{u}$

$$\hat{\mathbf{u}} = (\mathbf{I} - \mathbf{H})\mathbf{u}$$

- For instance,

$$\hat{u}_1 = (1 - h_{11})u_1 - \sum_{i=2}^{n} h_{1i}u_i$$

- Note that the residual is a function of all of the errors

Distribution of the residuals

$$\mathbb{E}[\hat{\mathbf{u}}|\mathbf{X}] = (\mathbf{I} - \mathbf{H})\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbf{0}$$

$$\text{Var}[\mathbf{u}|\mathbf{X}] = \sigma_u^2(\mathbf{I} - \mathbf{H})$$

- Variance of the ith residual:

$$\text{Var}[\hat{u}_i] = \sigma_u^2(1 - h_{ii})$$

- The residuals are not independent:
  - We know that the must sum up to 0: $\sum_i \hat{u}_i = 0$, so if I know $n - 1$ of them, I know the last one too.
- Residuals not independent, nor identically distributed, even when all the OLS assumptions hold
- Can't use them yet for checking Normality

Standardized residuals

- Problem: each residual has a different variance

$$\text{Var}[\hat{u}_i] = \sigma_u^2(1 - h_{ii})$$

- Possible solution: calculate standardized residuals by dividing by their variance:

$$\hat{u}_i' = \frac{\hat{u}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}$$

- Remember that $\mathbf{H} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$, which is observed
- Problem: $\hat{\sigma}^2$ depends on estimated residuals $\hat{u}_i$
- Numerator and denominator not independent $\rightsquigarrow$ unknown distribution

Studentized residuals

- Solution: estimate residual variance without residual i:

$$\widehat{\sigma}^2_{-i} = \frac{\mathbf{u}'\mathbf{u} - u_i^2/(1-h_{ii})}{n-k-2}$$

- Use this i-free estimate to standardize, which creates the studentized residuals:

$$\hat{u}_i^* = \frac{\hat{u}_i}{\widehat{\sigma}_{-i}\sqrt{1-h_{ii}}}$$

- If the errors are Normal, the studentized residuals follow a t distribution with $(n-k-2)$ degrees of freedom.
- Deviations from $t \implies$ violation of Normality

Buchanan vote example

```
mod <- lm(edaybuchanan ~ edaytotal, data = flvote)
summary(mod)
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 54.22945   49.14146    1.10     0.27
## edaytotal    0.00232    0.00031    7.48  2.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 333 on 65 degrees of freedom
## Multiple R-squared:  0.463,  Adjusted R-squared:  0.455
## F-statistic:   56 on 1 and 65 DF,  p-value: 2.42e-10
```

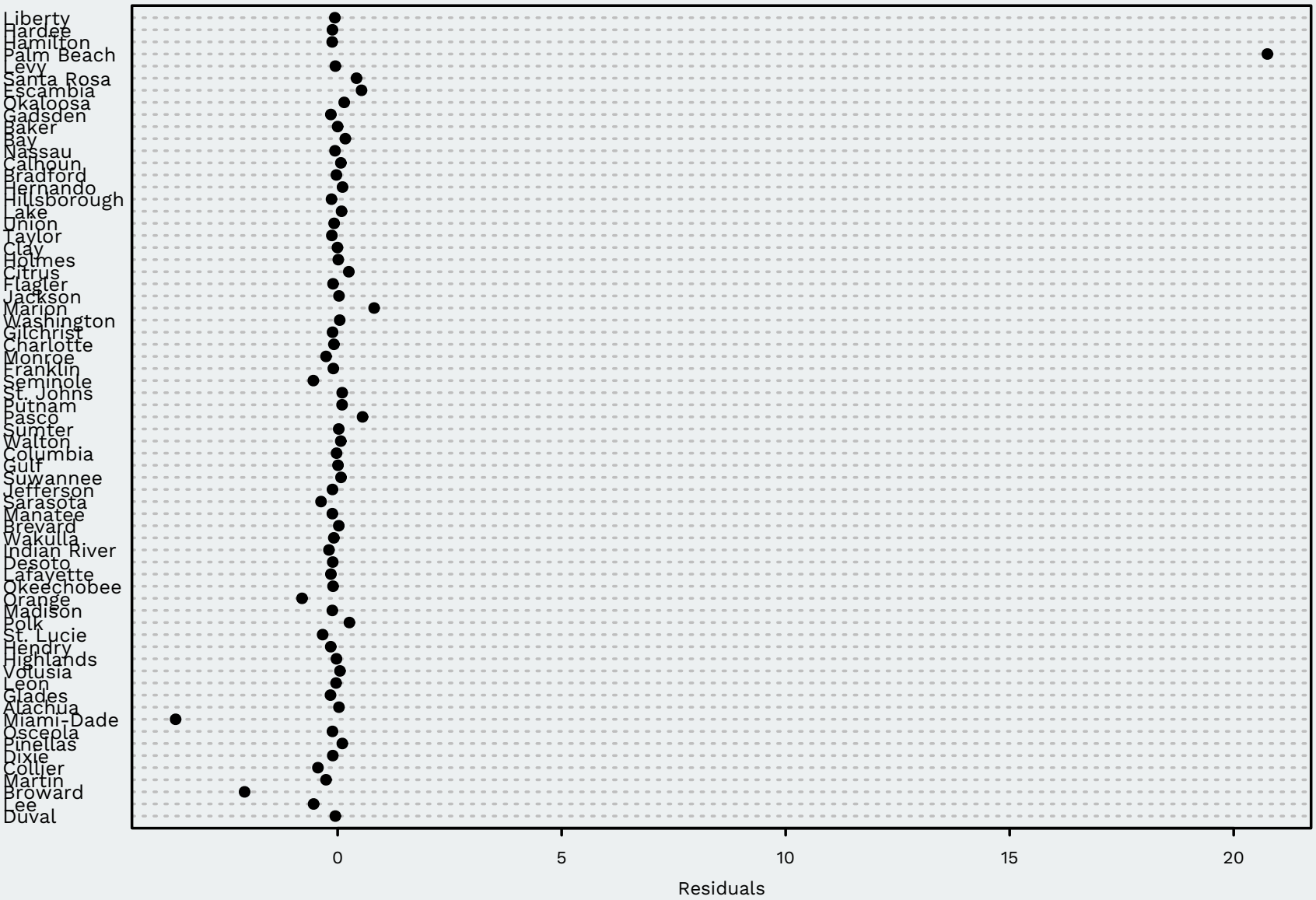
Buchanan residuals

```
resids <- residuals(mod)
stand.resids <- rstandard(mod)
student.resids <- rstudent(mod)
head(cbind(resids, stand.resids, student.resids))
```

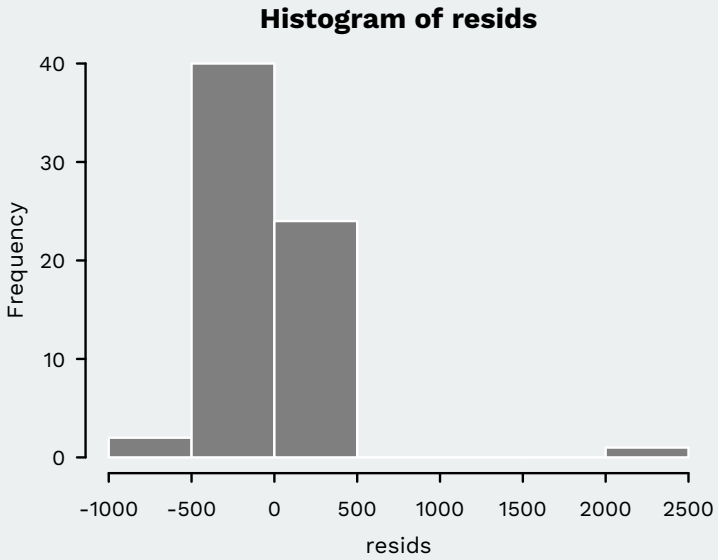
```
##    resids stand.resids student.resids
## 1  -16.94     -0.05201       -0.05161
## 2 -177.51     -0.53971       -0.53675
## 3 -595.20     -2.02639       -2.07743
## 4  -86.28     -0.26135       -0.25947
## 5 -146.31     -0.44306       -0.44030
## 6  -36.07     -0.10957       -0.10873
```

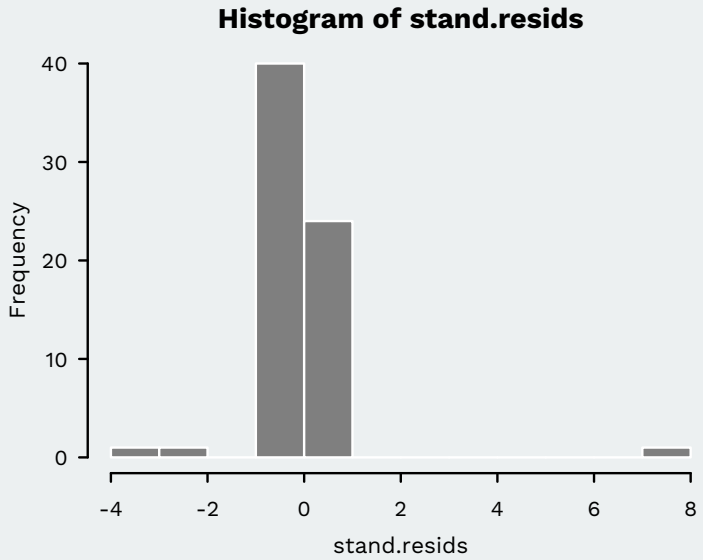
```
dotchart(student.resids, flvote$county, xlab = "Residuals")
```

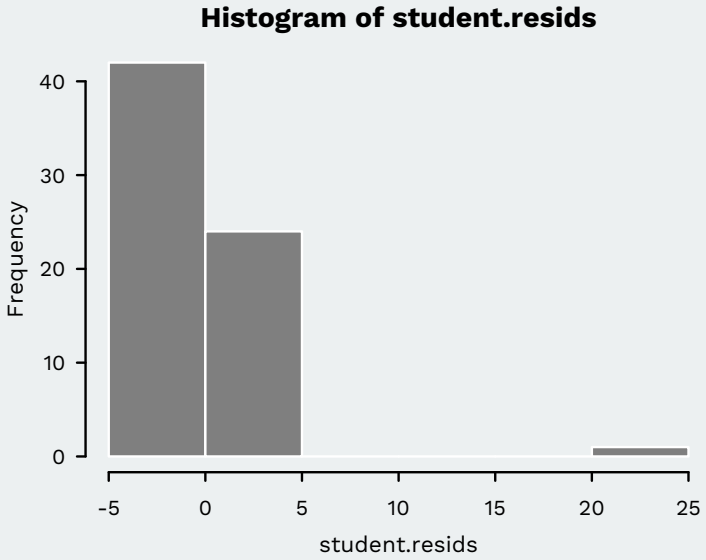
Plotting the residuals



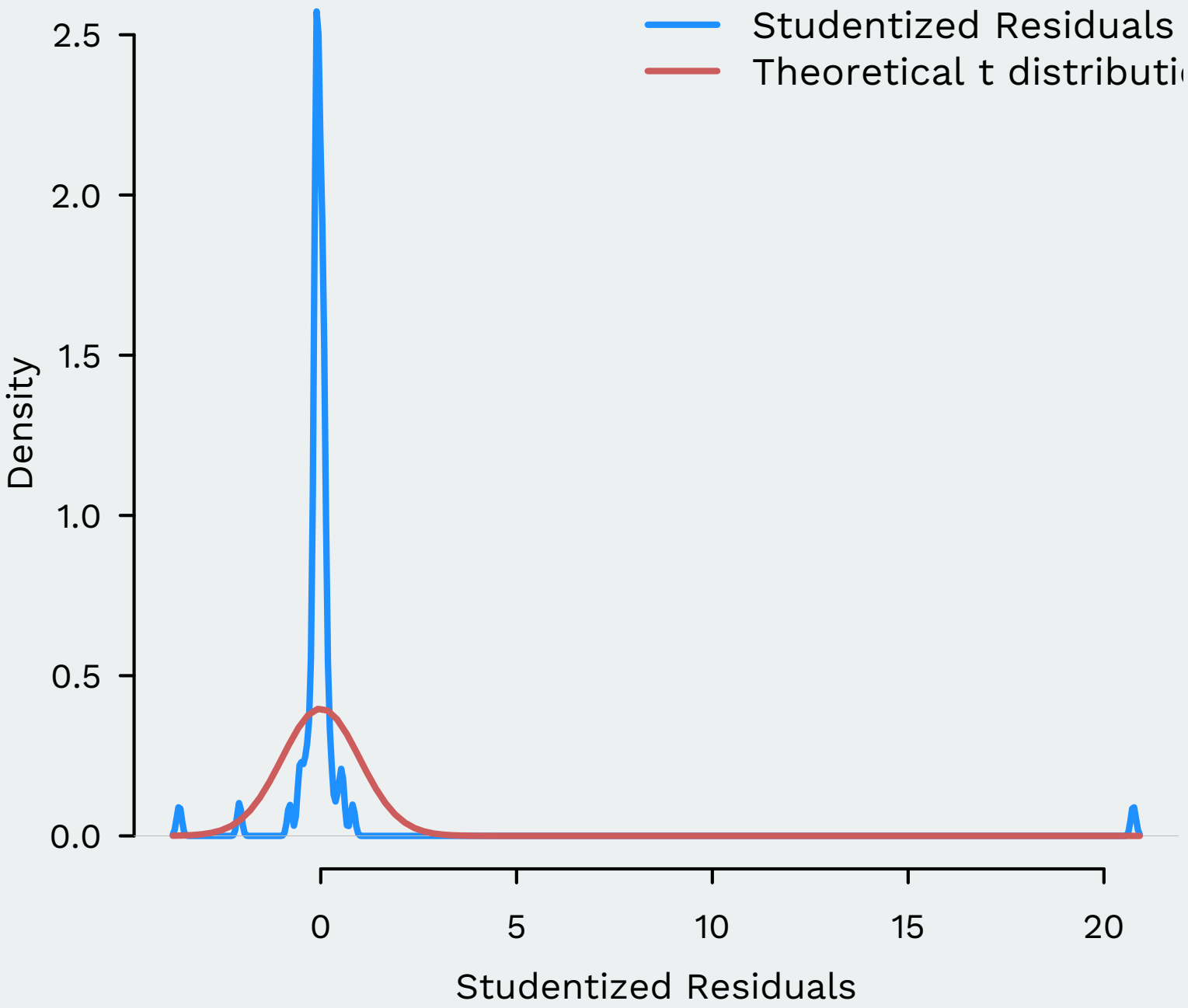
Plotting the residuals







Plotting the residuals

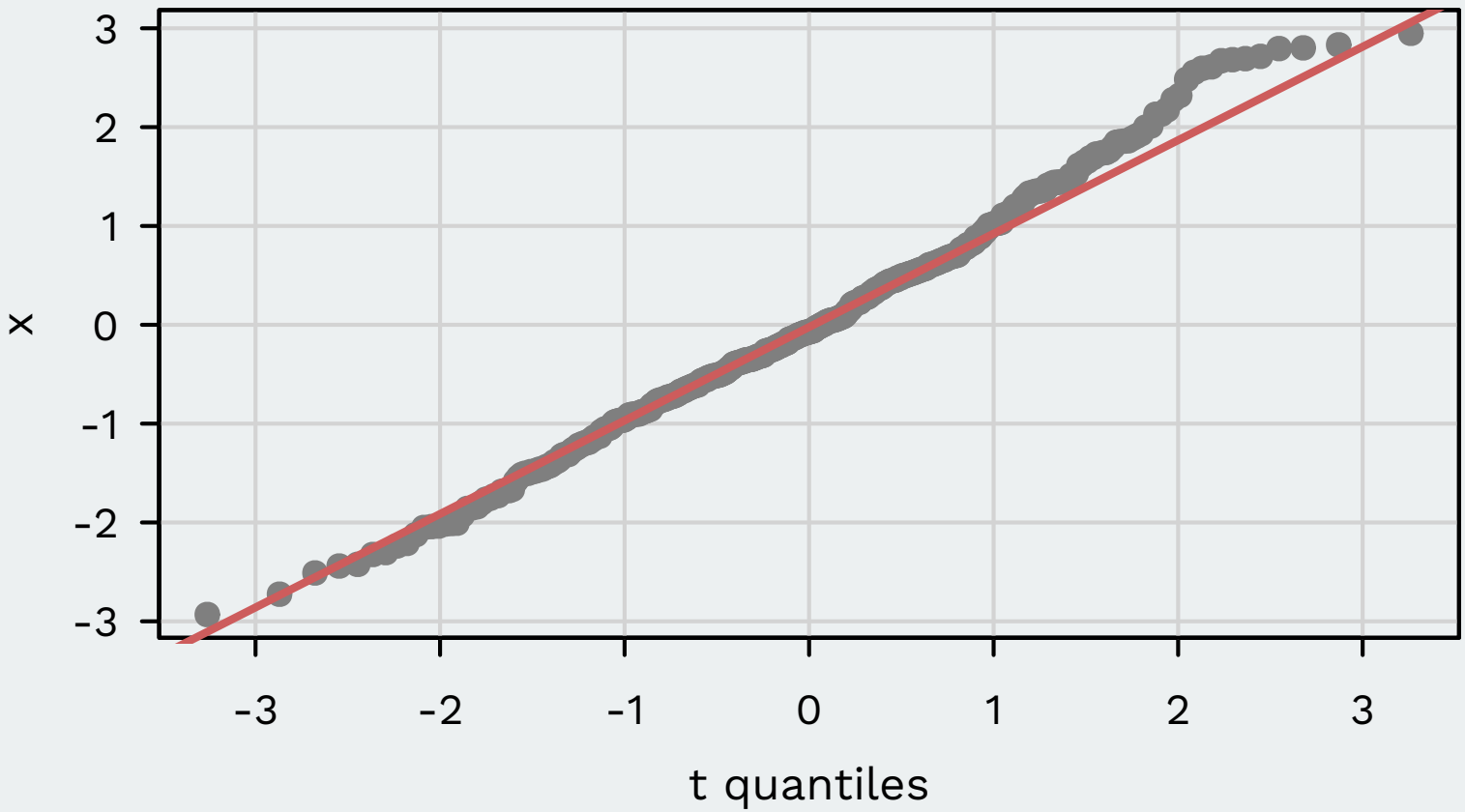


Quantile-Quantile plots

- Quantile-quantile plot or QQ-plot is useful for comparing distributions
- Plots the quantiles of one distribution against those of another distribution
- For example, one point is the (m_x, m_y) where m_x is the median of the x distribution and m_y is the median for the y distribution
- If distributions are equal $\implies$ 45 degree line

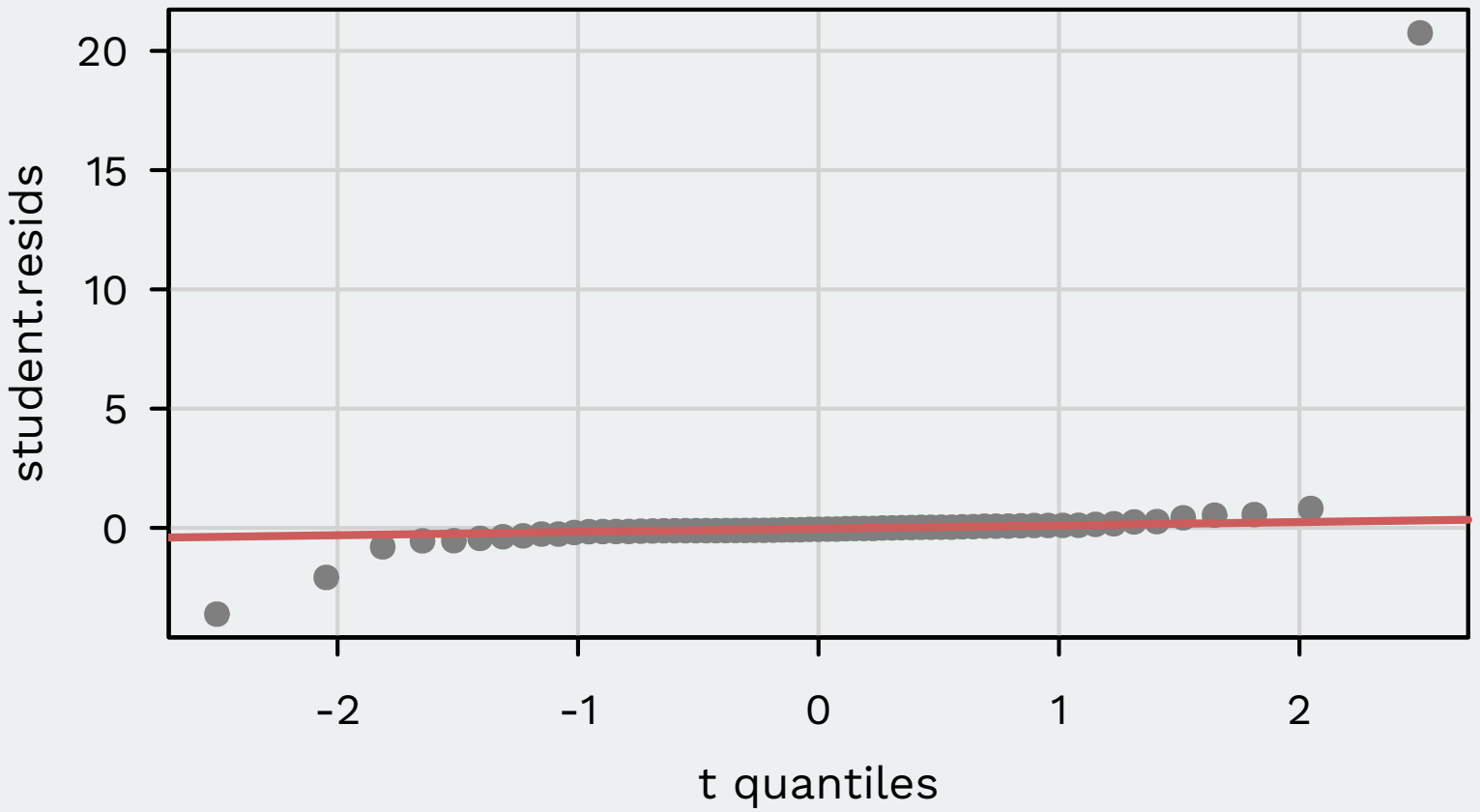
Good QQ-plot

```
library(car)
x <- rt(500, df = 50)
qqPlot(x, distribution = "t", df = 50, pch = 19, cex = 1, col = "gr
    col.lines = "indianred", las = 1, envelope = TRUE)
```



Buchanan QQ-plot

```
qqPlot(student.resids, distribution = "t", df = nrow(flvote) -
    2, envelope = TRUE, pch = 19, cex = 1, col = "grey50", co
    las = 1)
```



Dealing with non-Normal errors

- Remove problematic observations (be transparent!)
- Add or drop variables in $\mathbf{X}$
- Transform $\mathbf{y}$ ($\log(\mathbf{y})$)

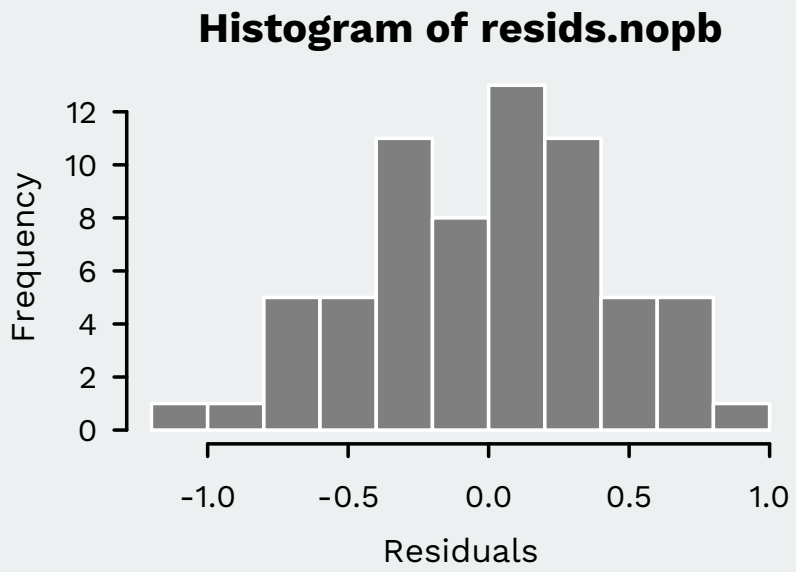
Buchanan revisited

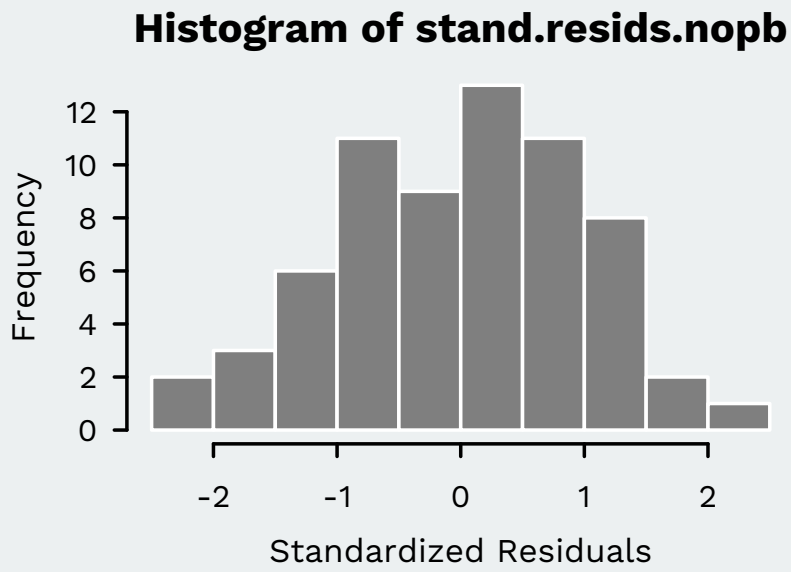
```
flvote.nopb <- flvote[flvote$county != "Palm Beach", ]
mod.nopb <- lm(log(edaybuchanan) ~ log(edaytotal), data = flvote.nopb)
resids.nopb <- residuals(mod.nopb)
stand.resids.nopb <- rstandard(mod.nopb)
student.resids.nopb <- rstudent(mod.nopb)
summary(mod.nopb)
```

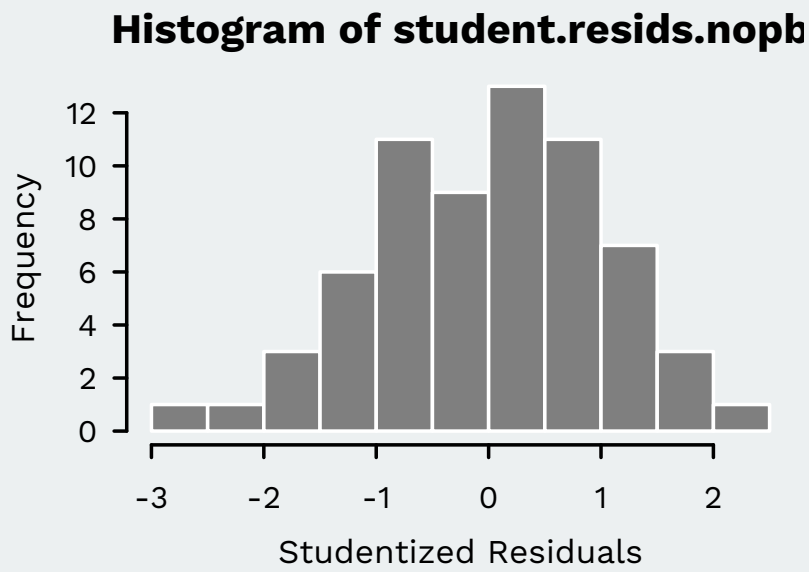
```
clipped.print(summary(mod.nopb))
```

```
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -2.4860     0.3789   -6.56  1.1e-08 ***
## log(edaytotal)   0.7031     0.0362   19.42  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.436 on 64 degrees of freedom
## Multiple R-squared:  0.855,  Adjusted R-squared:  0.853
## F-statistic:  377 on 1 and 64 DF,  p-value: <2e-16
```

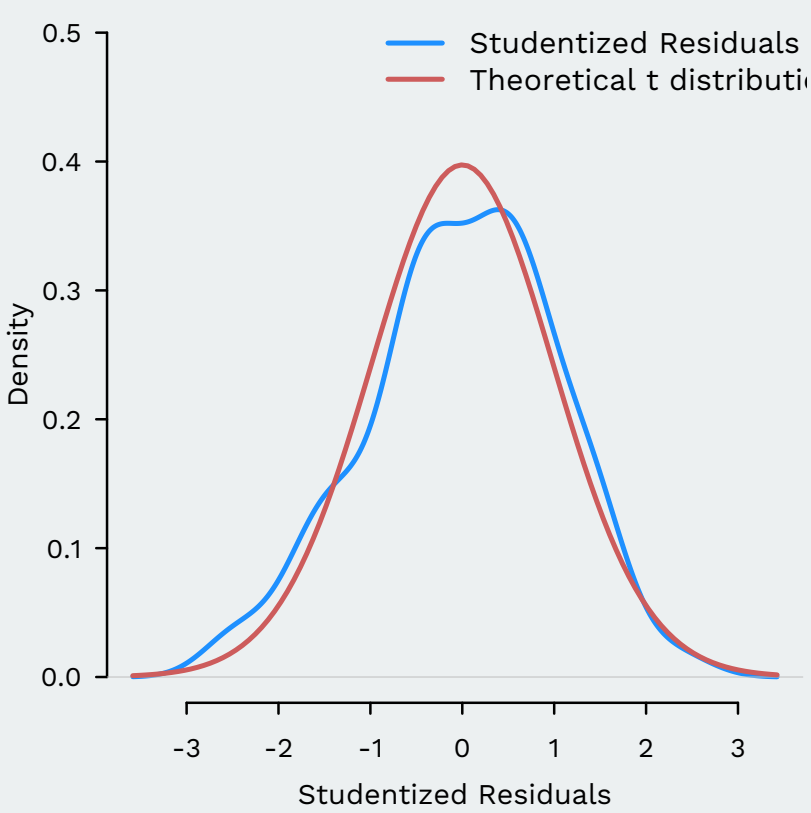
Buchanan revisited

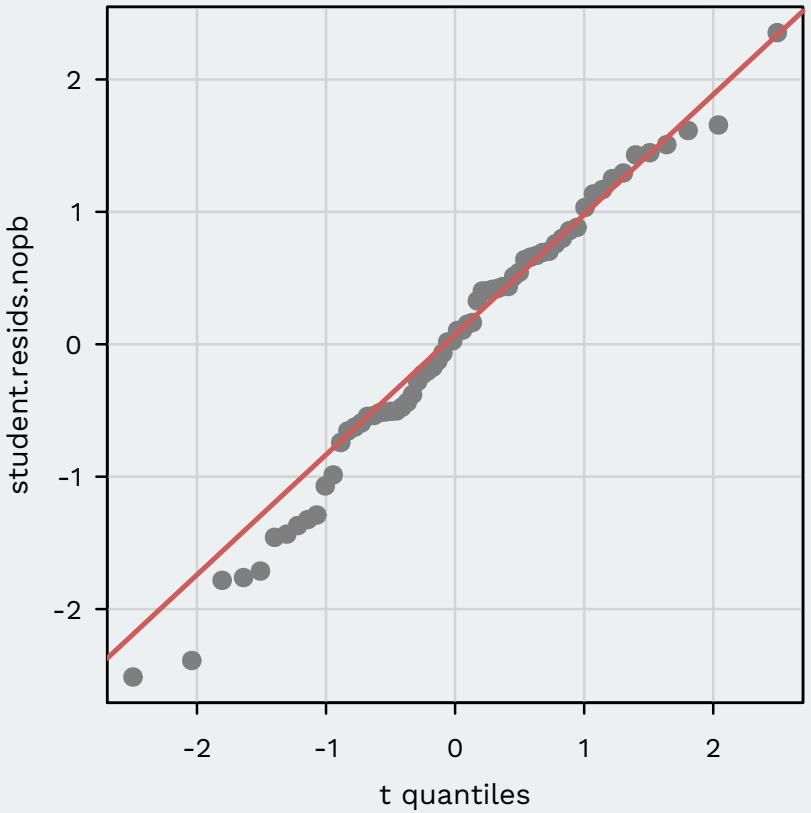






Buchanan revisited





2/ Nonlinearity of the regression function

Buchanan model, part 2

```
mod3 <- lm(edaybuchanan ~ edaytotal + absnbuchanan, data = fl
summary(mod3)
```

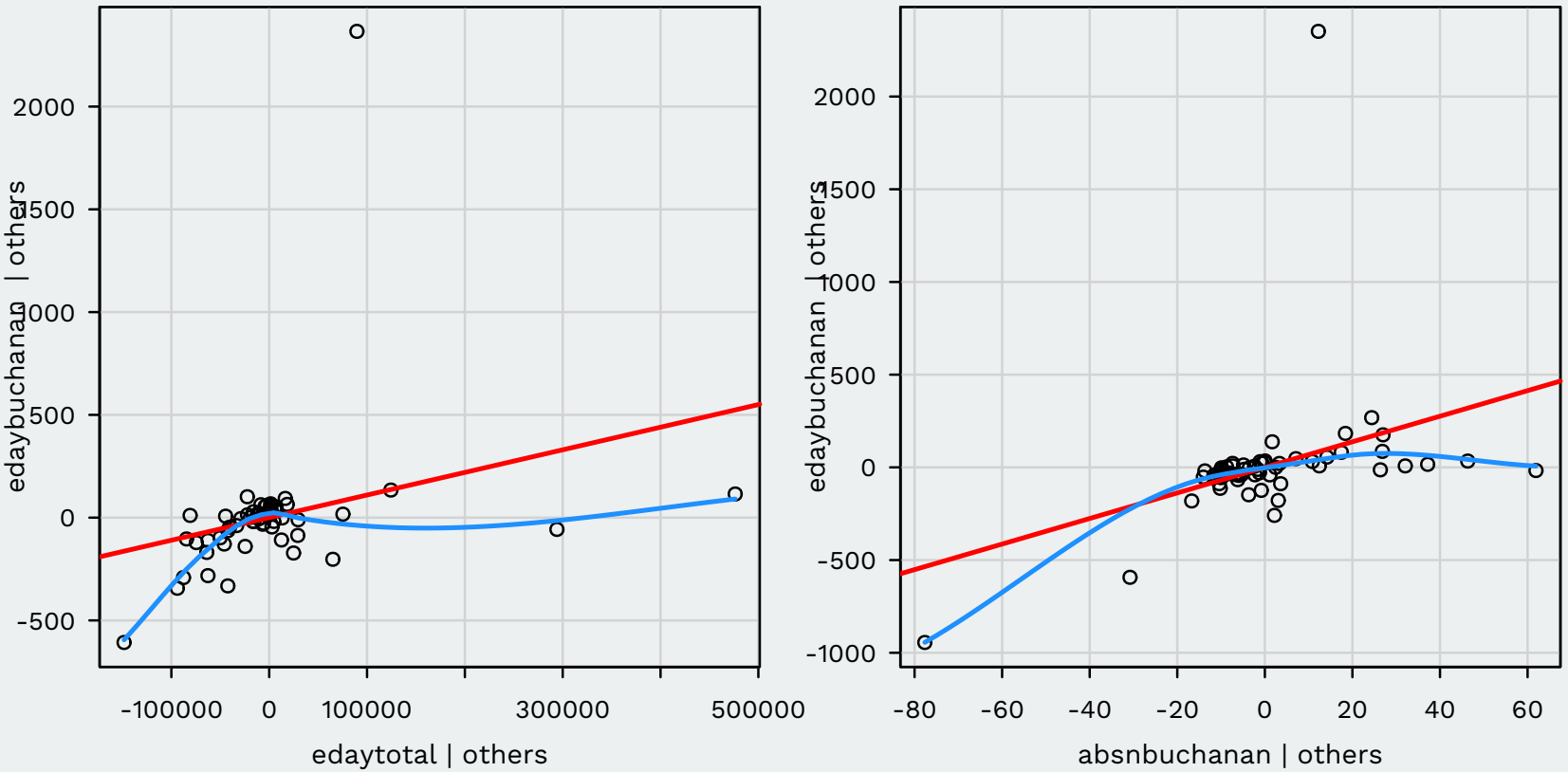
```
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -29.34807   55.19635   -0.53   0.5969
## edaytotal      0.00110    0.00048    2.29   0.0253 *
## absnbuchanan   6.89546    2.12942    3.24   0.0019 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 317 on 61 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.536,  Adjusted R-squared:  0.521
## F-statistic: 35.2 on 2 and 61 DF,  p-value: 6.71e-11
```

Added variable plot

- Need a way to visualize conditional relationship between Y and X_j
- How to construct an added variable plot:
  1. Get residuals from regression of Y on all covariates except X_j
  2. Get residuals from regression of X_j on all other covariates
  3. Plot residuals from (1) against residuals from (2)
- In R: `avPlots(model)` from the `car` package
- OLS fit to this plot will have exactly $\widehat{\beta}_j$ and 0 intercept
- Use local smoother (loess) to detect any non-linearity

Buchanan AV plot

```
par(mfrow = c(1, 2))
out <- avPlots(mod3, "edaytotal")
lines(loess.smooth(x = out$edaytotal[, 1], y = out$edaytotal[, 2]),
    col = "dodgerblue", lwd = 2)
out2 <- avPlots(mod3, "absnbuchanan")
lines(loess.smooth(x = out2$absnbuchanan[, 1], y = out2$absnbuchanan[,
    2]), col = "dodgerblue", lwd = 2)
```



How to deal with non-linearity

- Breaking up categorical variables into dummy variables
- Including interaction terms
- Including polynomial terms
- Using transformations
- Using more flexible models:
  - Generalized additive models and splines allow the data to tell us what the functional form is.
  - Complicated math, but important ideas.

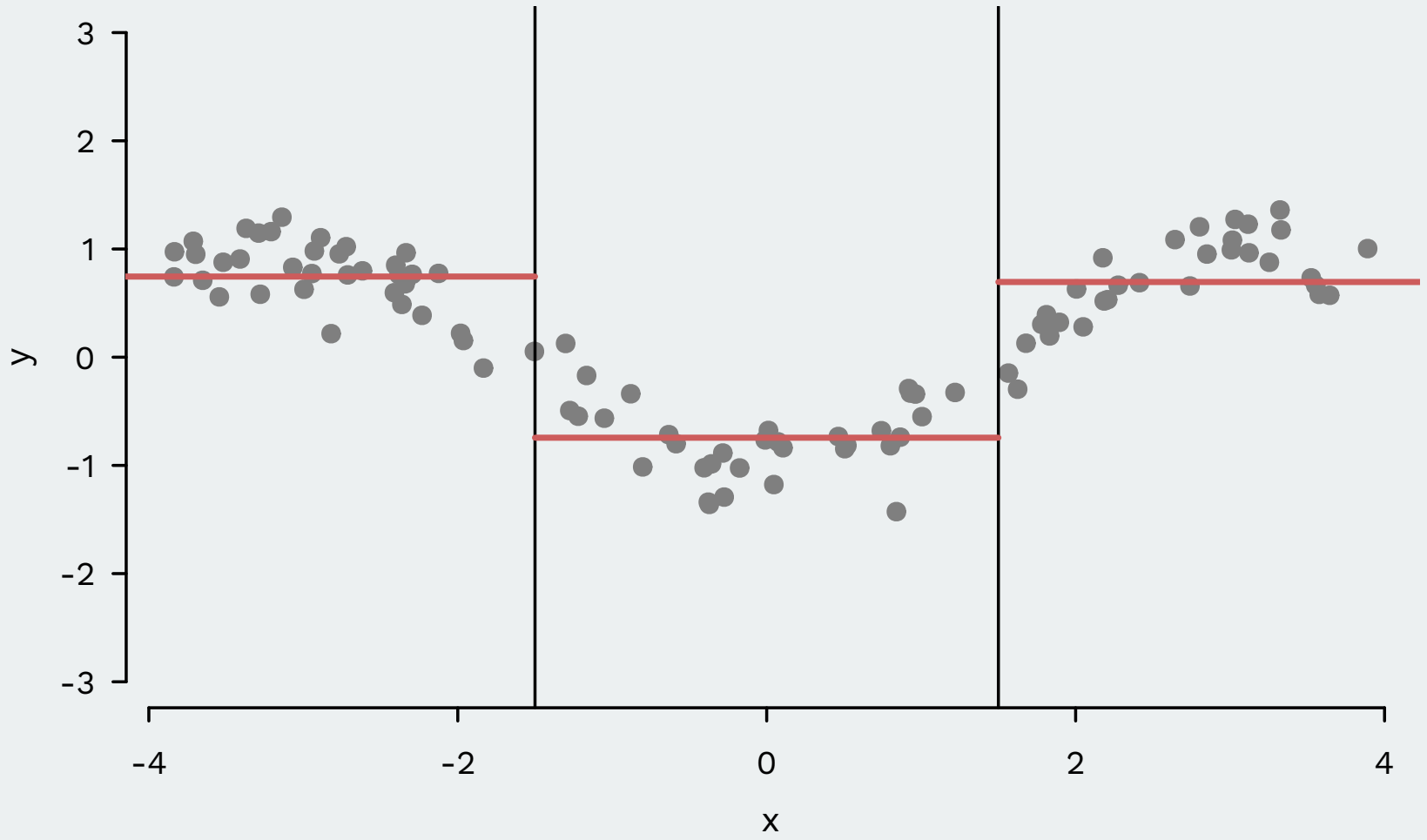
Basis functions

- Basis functions are the function of X_i that we include in the model:
  - Examples we've seen: $h_m(X_i) = X_i$, $h_m(X_i) = X_i^2$, $h_m(X_i) = \log(X_i)$
- Different basis functions will allow for different forms of non-linearity
- We could always break up X_i into bins and estimate piecewise constant:

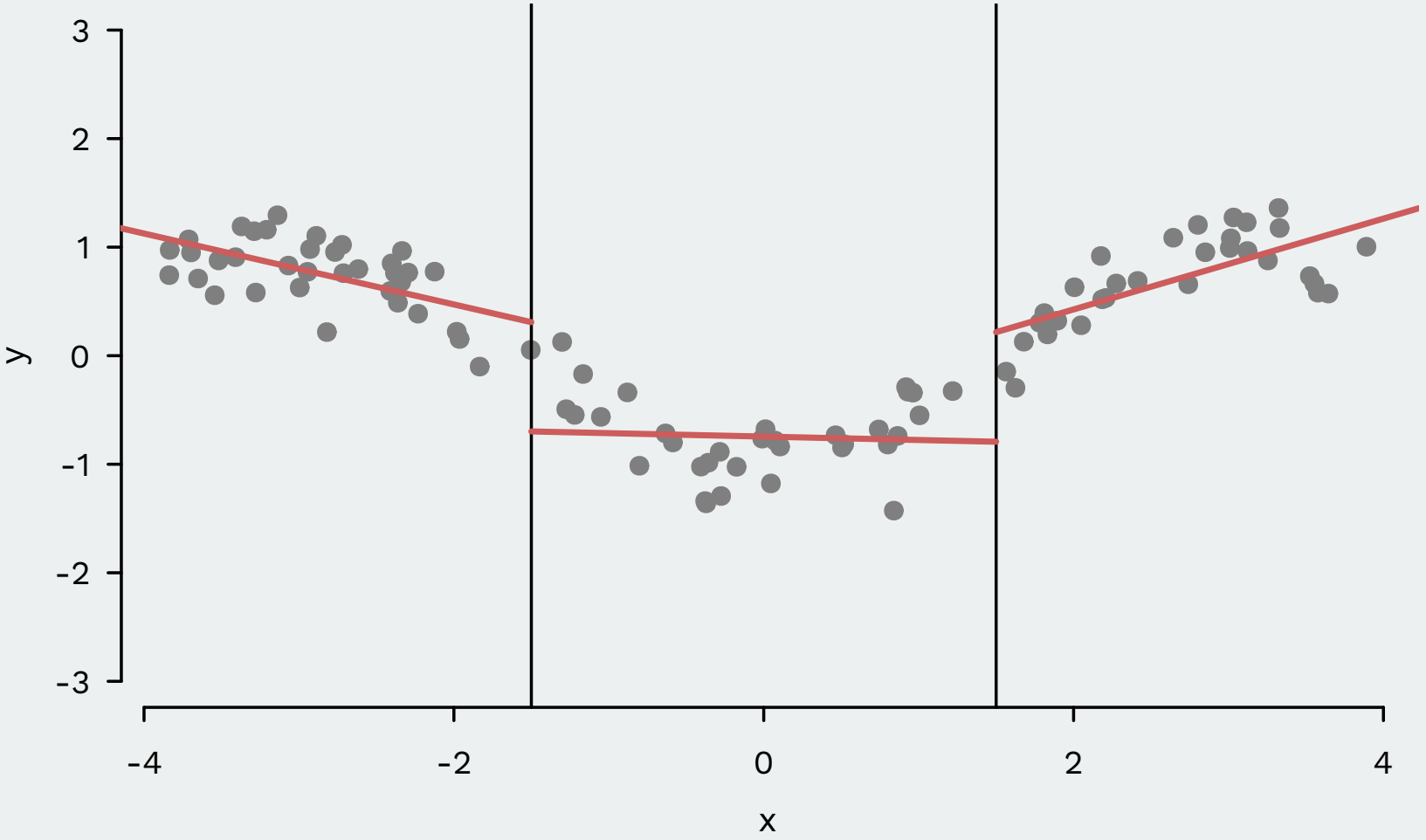
$$h_1 = \mathbb{1}(X_i < b_1), \quad h_2 = \mathbb{1}(b_1 < X_i < b_2), \quad h_3 = \mathbb{1}(X_i > b_2)$$

- $b_1 < b_2$ are knots

Piecewise constant



Piecewise linear



Continuous piecewise linear

- Problem: piecewise functions are discontinuous.
- Can use clever basis functions to get continuous piecewise linear function of X_i:

$$h_1(X_i) = 1, \qquad h_2(X_i) = X_i,$$
$$h_3(X_i) = (X_i - b_1)_+, \qquad h_4(X_i) = (X_i - b_2)_+$$

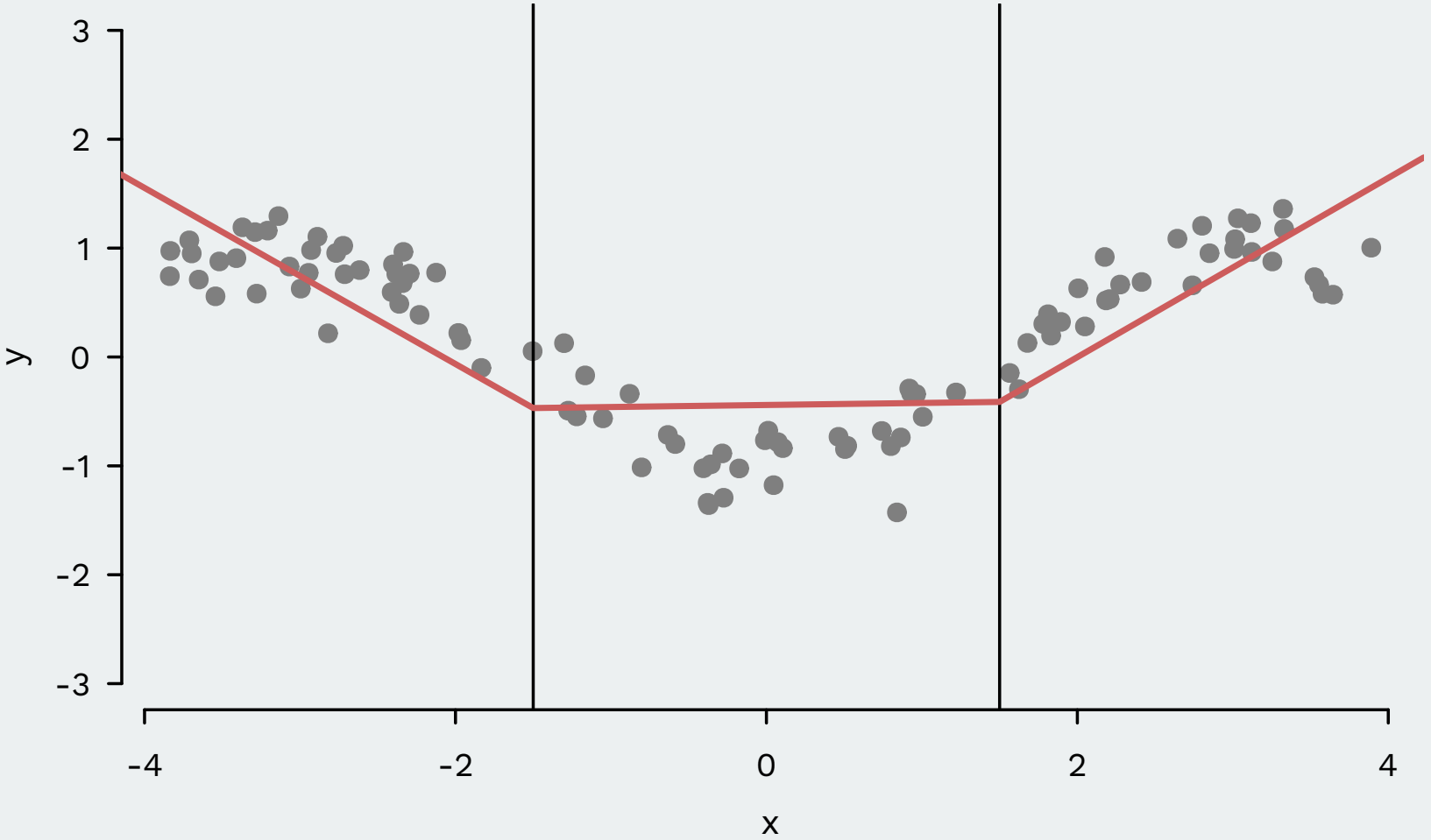
- $(X_i - b_1)_+ = X_i - b_1$ when $X_i > b_1$, 0, otherwise
- Regression with these basis functions:

$$Y_i = \beta_1 h_1(X_i) + \beta_2 h_2(X_i) + \beta_3 h_3(X_i) + \beta_4 h_4(X_i) + \varepsilon_i$$

  - $\beta_2 =$ slope when $X_i < b_1$
  - $\beta_2 + \beta_3 =$ slope when $b_1 < X_i < b_2$
  - $\beta_2 + \beta_3 + \beta_4 =$ slope when $X_i > b_2$

Continuous piecewise linear

```
h2 <- x
h3 <- 1 * (x > -1.5) * (x - -1.5)
h4 <- 1 * (x > 1.5) * (x - 1.5)
reg <- lm(y ~ h2 + h3 + h4)
```



Cubic splines

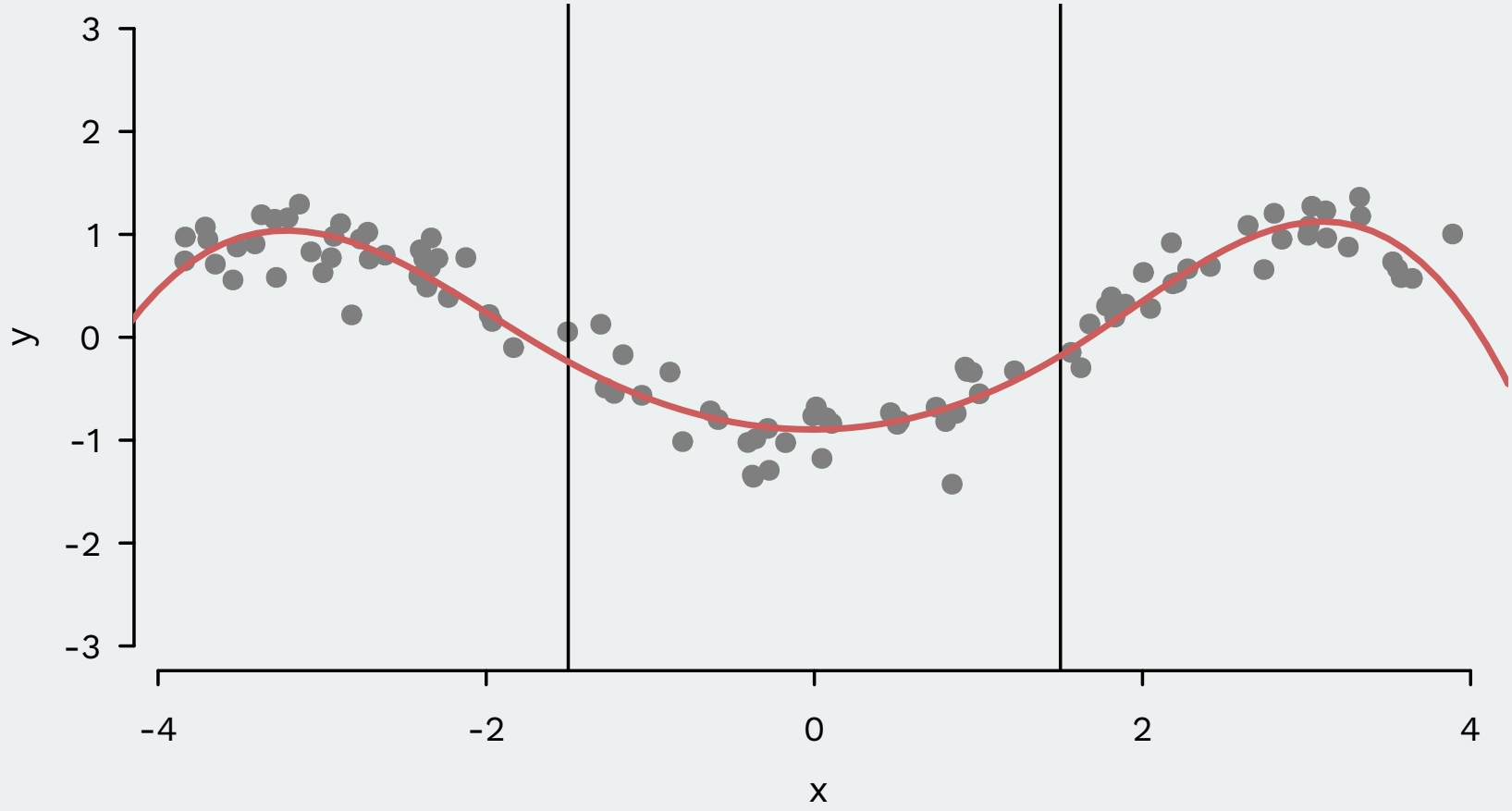
- Continuous piecewise linear has "kinks" at the knots, but we probably want "smooth" functions.
  - What does smooth mean? Continuous derivatives!
  - $\rightsquigarrow$ use higher-order polynomials in the basis functions
- Cubic spline basis: bases that produce continuous functions with continuous first and second derivatives

$$h_1(X_i) = 1, \qquad h_2(X_i) = X_i, \qquad h_3(X_i) = X_j^2$$
$$h_4(X_i) = X_i^3, \qquad h_5(X_i) = (X_i - b_1)_+^3, \qquad h_6(X_i) = (X_i - b_2)_+^3$$

- Basic idea: local polynomial regression (between knots) that have to connect and be smooth at the knots.

Cubic spline

```
h2 <- x
h3 <- x^2
h4 <- x^3
h5 <- 1 * (x > -1.5) * (x - -1.5)^3
h6 <- 1 * (x > 1.5) * (x - 1.5)^3
reg <- lm(y ~ h2 + h3 + h4 + h5 + h6)
```



Knotty problems

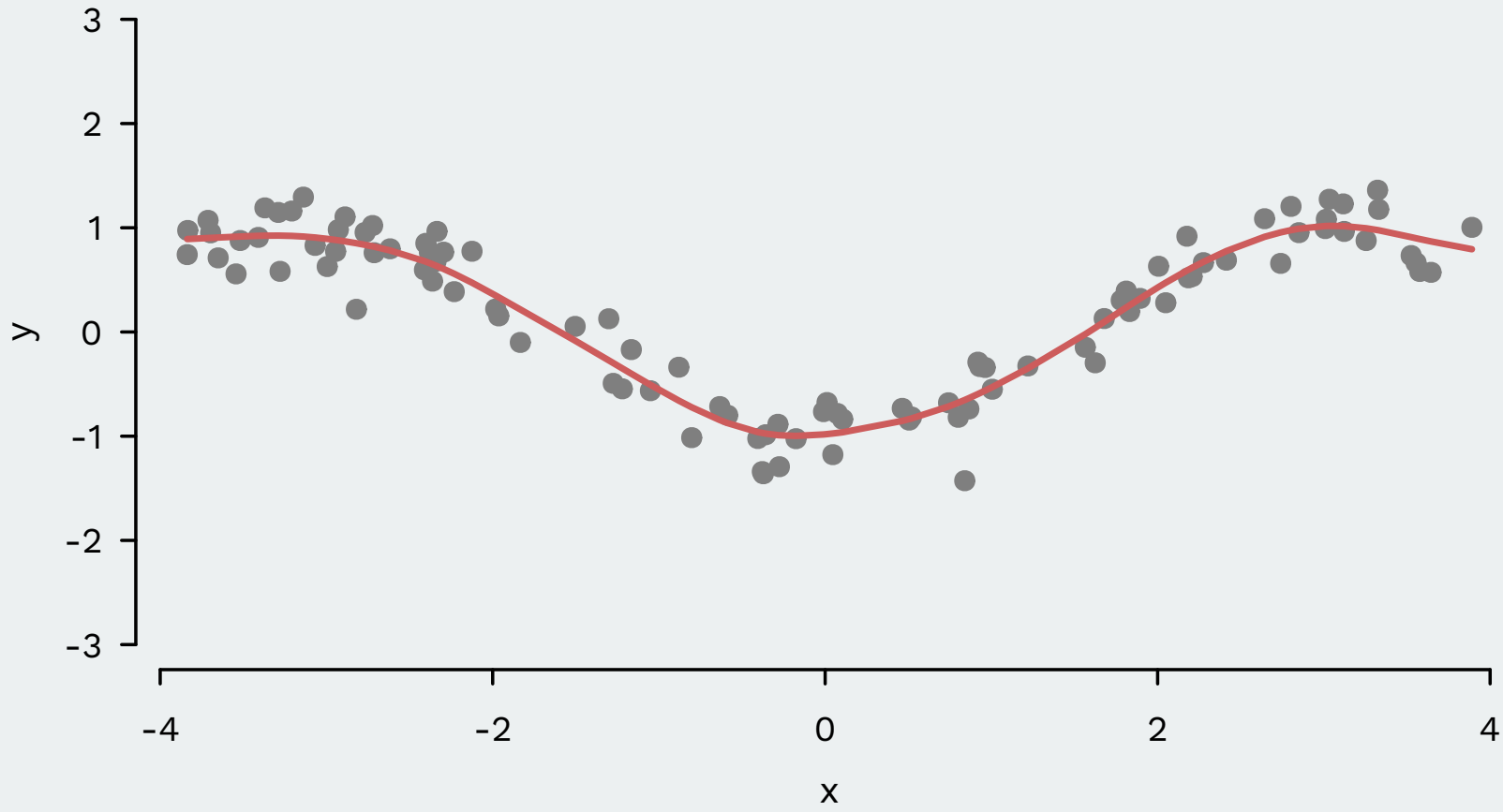
- Any function can be approximated as we increase the number of knot points.
- How to choose the number/location of knot points?
  - More knot points $\rightsquigarrow$ "rougher" function, less in-sample bias, more variance.
  - Fewer knot points $\rightsquigarrow$ "smoother" function, more in-sample bias, less variance.
- In-sample fit might be great, out-of-sample fit might be terrible.

Cross-validation

- General strategy for bias-variance trade-offs: cross-validation.
- Set aside units to test out-of-sample prediction
- Cross-validation procedure:
  1. Choose a number of evenly spread knots, b.
  2. Withhold unit i, estimate the CEF of y_i given X_i using a cubic spline with b knots.
  3. Get predicted value for i, $\hat{y}_{ib}^{-i}$ and caculate squared prediction error: $(y_i - \hat{y}_{ib}^{-i})^2$.
  4. Repeat 2-3 for each observation and take that average to get the MSE with b knots.
  5. Repeat 1-4 for different values of b and choose the value of b that has the lowest MSE.

Automatic knot selection

```
smth <- smooth.spline(x, y)
plot(x, y, ylim = c(-3, 3), pch = 19, col = "grey50", bty = "n")
lines(smth, col = "indianred", lwd = 2)
```



Generalized additive models

- Generalized additive models (GAMs) allow you to estimate the spline of any particular variable in the regression.
  - Each spline is additive: $y_i = f_1(X_1) + f_x(X_2) + \varepsilon_i$
- Can plot the AV-plot of the spline to get a sense for the nonlinearity of the functional form.
- Use cross-validation to select the number of knots

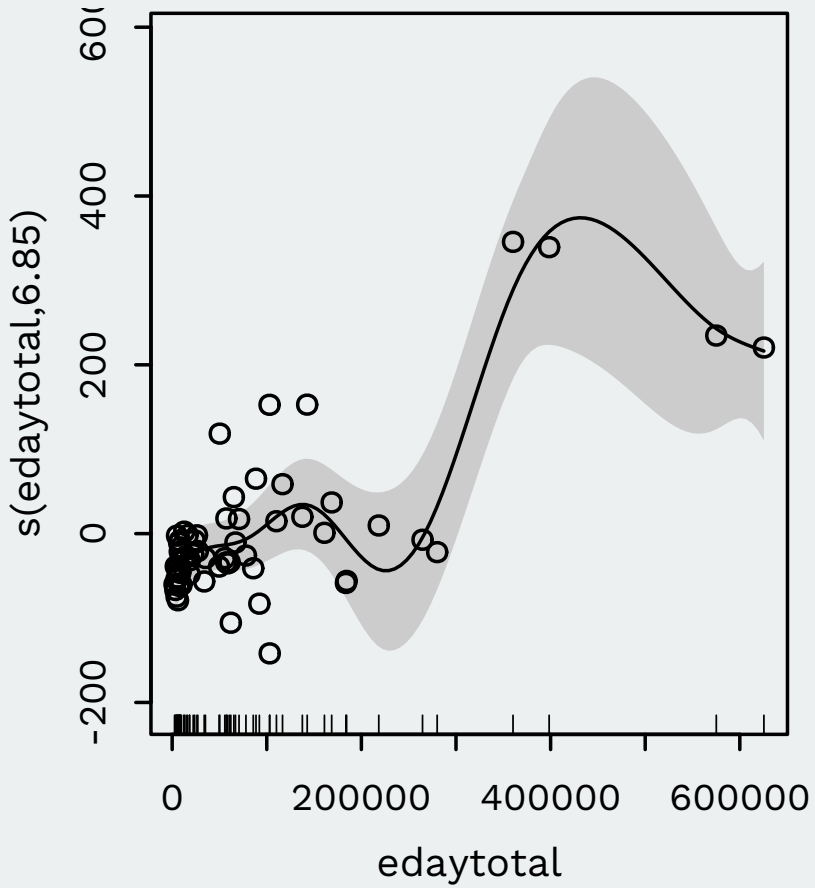
GAM example fit

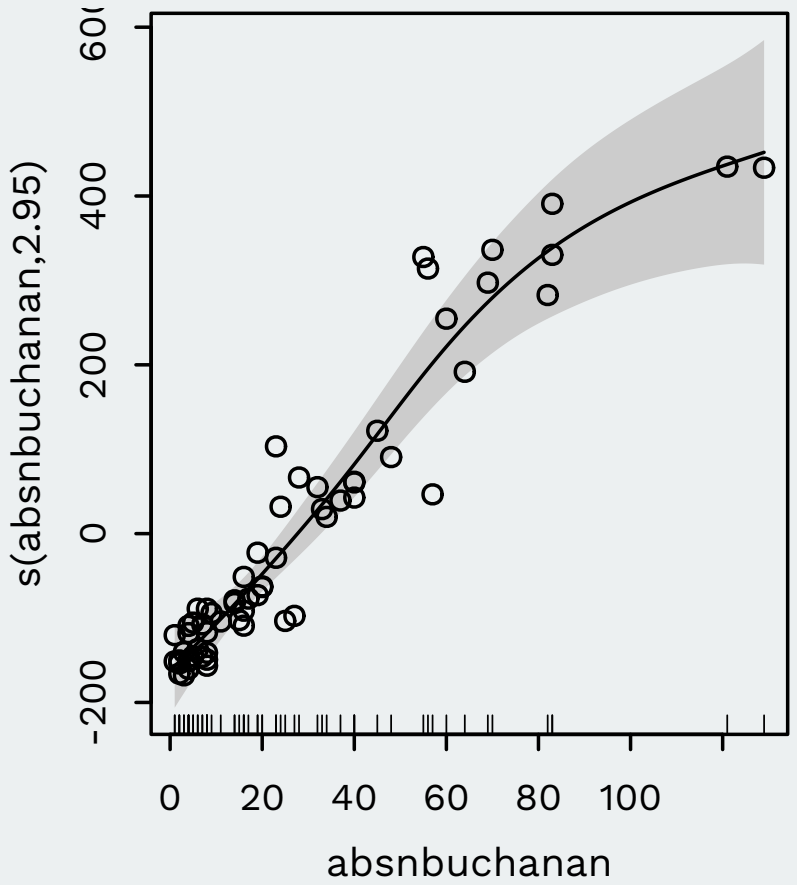
```
## library(mgcv)  ## GAM package
out <- gam(edaybuchanan ~ s(edaytotal) + s(absnbuchanan), data = flvote,
    subset = county != "Palm Beach")
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## edaybuchanan ~ s(edaytotal) + s(absnbuchanan)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   221.84       6.41    34.6   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                  edf Ref.df    F p-value
## s(edaytotal)    6.85   7.82 10.6 1.6e-09 ***
## s(absnbuchanan) 2.95   3.64 22.6 1.6e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =   0.95   Deviance explained = 95.8%
## GCV =   3129  Scale est. = 2592.3    n = 63
```

Example: generalized additive models

```
plot(out, shade = TRUE, residual = TRUE, pch = 1)
```





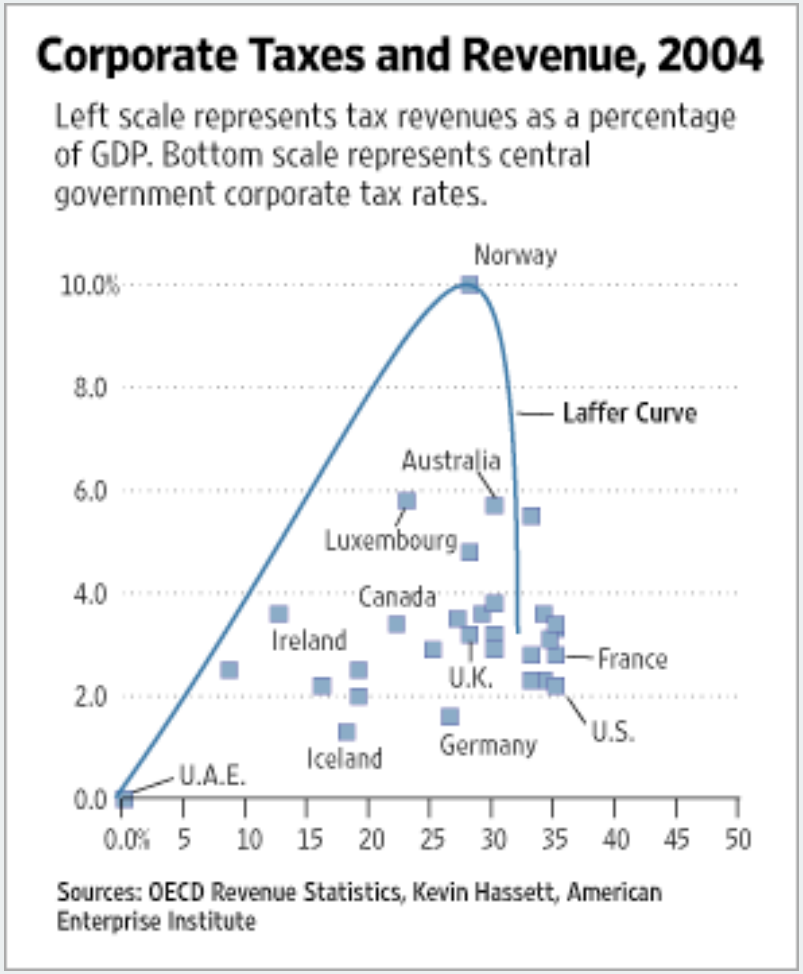
3/ Outliers, leverage points, and influential observations

The trouble with Norway

- Lange and Garrett (1985): organizational and political power of labor interact to improve economic growth
- Jackman (1987): relationship just due to North Sea Oil?
- Table guide:
  - x_1 = organizational power of labor
  - x_2 = political power of labor
  - Parentheses contain t-statistics

	Constant	x_1	x_2	$x_1 \cdot x_2$
Norway Obs Included	.814	-.192	-.278	.137
	(4.7)	(2.0)	(2.4)	(2.9)
Norway Obs Excluded	.641	-.068	-.138	.054
	(4.8)	(0.9)	(1.5)	(1.3)

Creative curve fitting with Norway

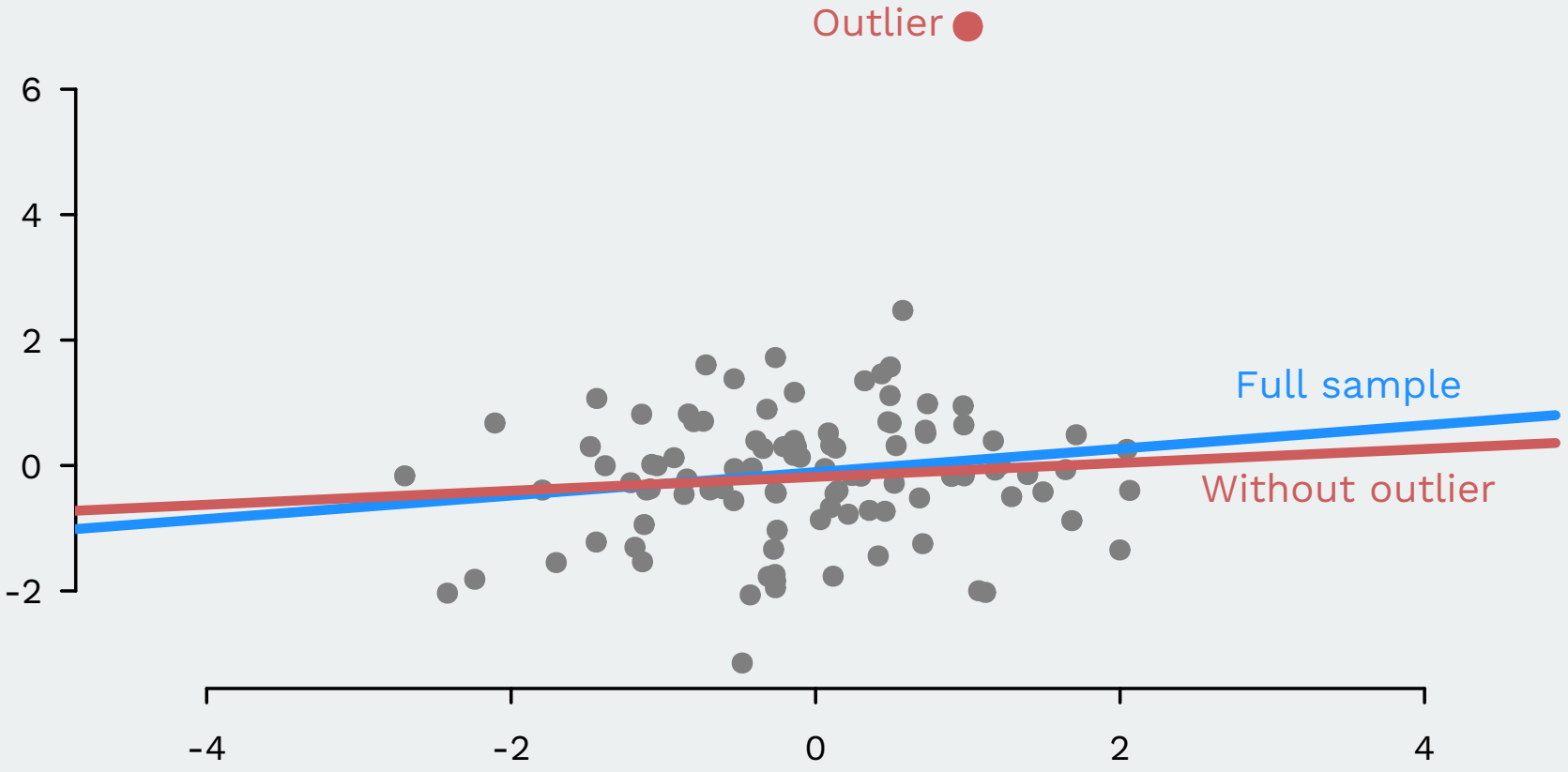


Three types of extreme values

1. Outlier: extreme in the y direction
2. Leverage point: extreme in one x direction
3. Influence point: extreme in both directions

- Not all of these are problematic
- If the data are truly "contaminated" (come from a different distribution), can cause inefficiency and possibly bias
- Can be a violation of iid (not identically distributed)

Outlier definition

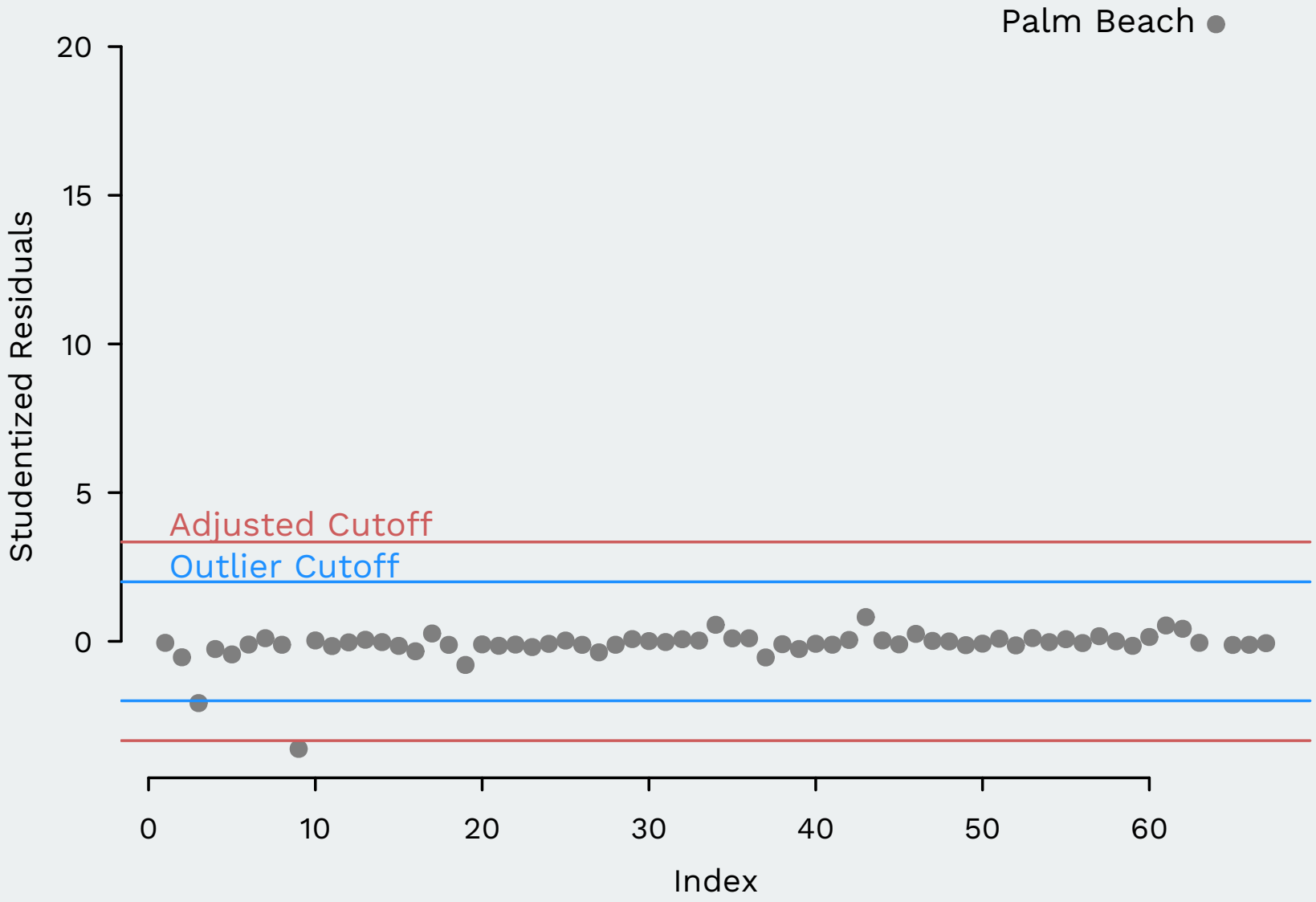


- An outlier is a data point with very large regression errors, u_i
- Very distant from the rest of the data in the y-dimension
- Increases standard errors (by increasing $\widehat{\sigma}^2$)
- No bias if typical in the x's

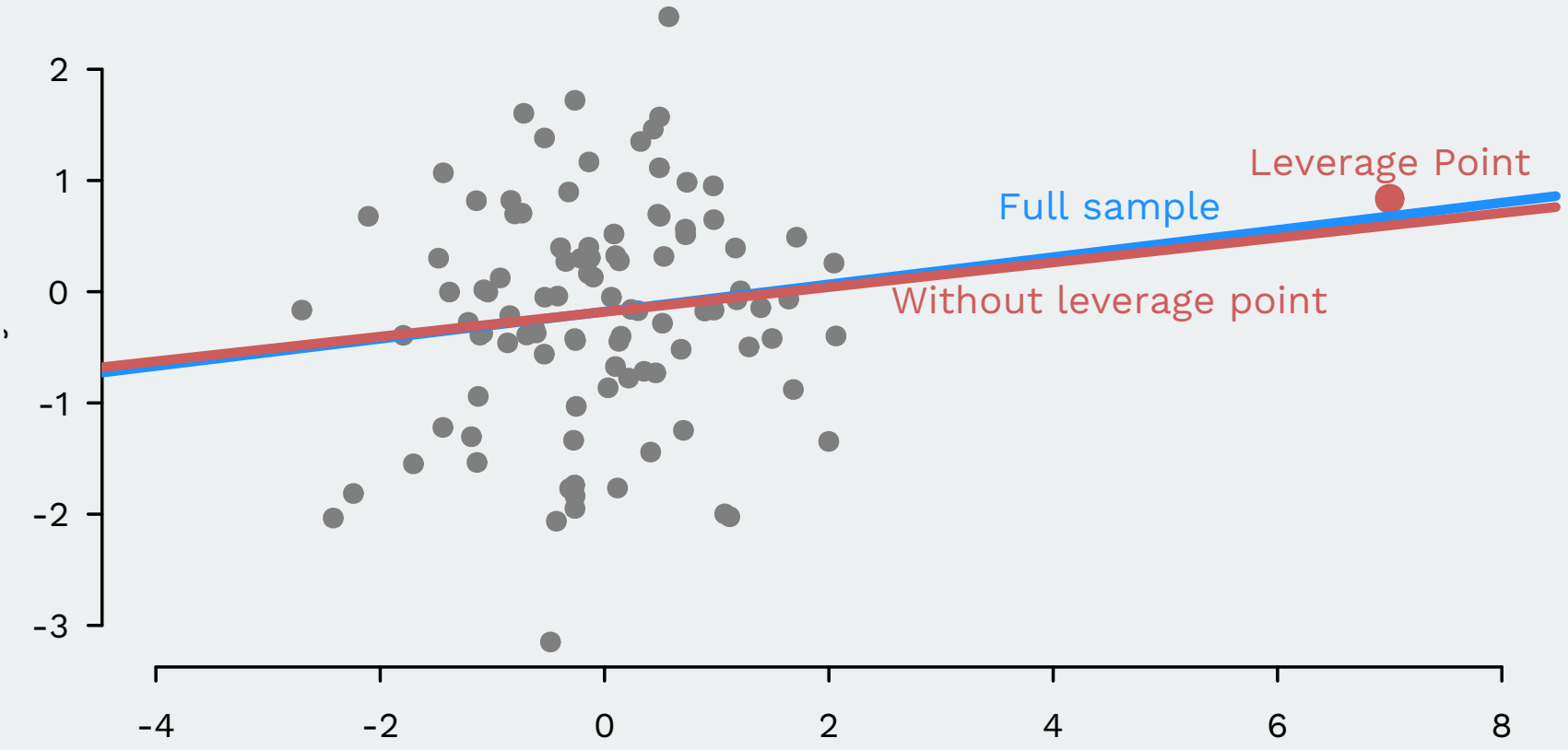
Detecting outliers

- Use studentized residuals, u_i^*, since outliers can skew the residual variance upward
  - $\widehat{\sigma}^2 >> \widehat{\sigma}^2_{-i}$ if i is an outlier
  - $u_i^* \sim t_{n-k-2}$, when $u_i \sim N(0, \sigma^2)$
- Rule of thumb: $|\widehat{u}_i^*| > 2$ will be relatively rare
- Extreme outliers, $|\widehat{u}_i^*| > 4 - 5$ are much less likely
- People usually adjust cutoff for multiple testing

Buchanan outliers



Leverage point definition



- Values that are extreme in the x direction
- That is, values far from the center of the covariate distribution
- Decrease SEs (more X variation)
- No bias if typical in y dimension

Hat values

$$\hat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y}$$

- For a particular observation i, we can show this means:

$$\hat{y}_i = \sum_{j=1}^{n} h_{ij} y_j$$

- h_{ij} = importance of observation j is for the fitted value $\hat{y}_i$
- Leverage/hat values: $h_i = h_{ii}$ diagonal entries of the hat matrix
- With a simple linear regression, we have

$$h_i = \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum_{j=1}^{n} (X_j - \overline{X})^2}$$

  - $\rightsquigarrow$ how far i is from the center of the $\mathbf{X}$ distribution

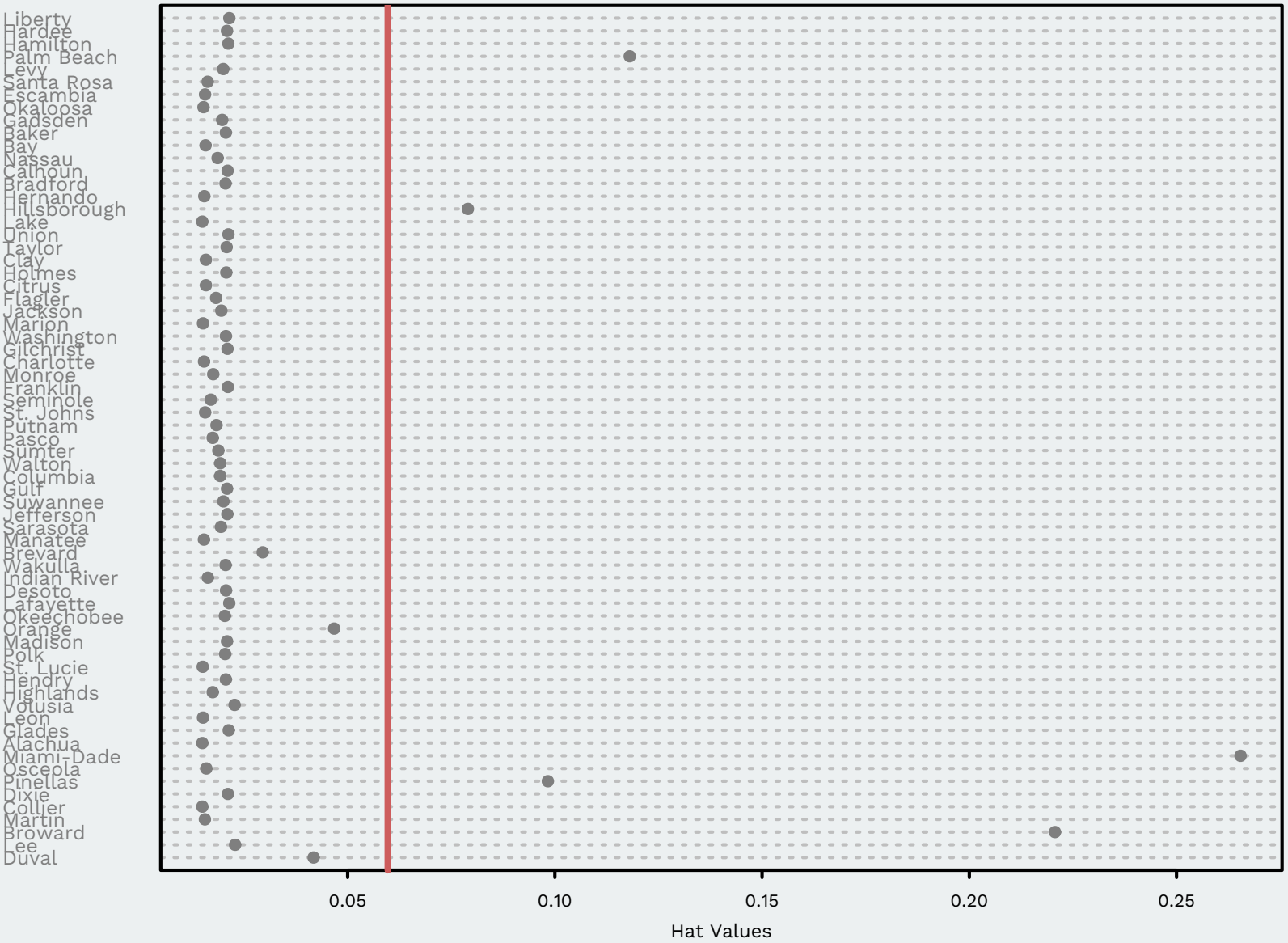
- Rule of thumb: examine hat values greater than $2(k+1)/n$

Buchanan hats

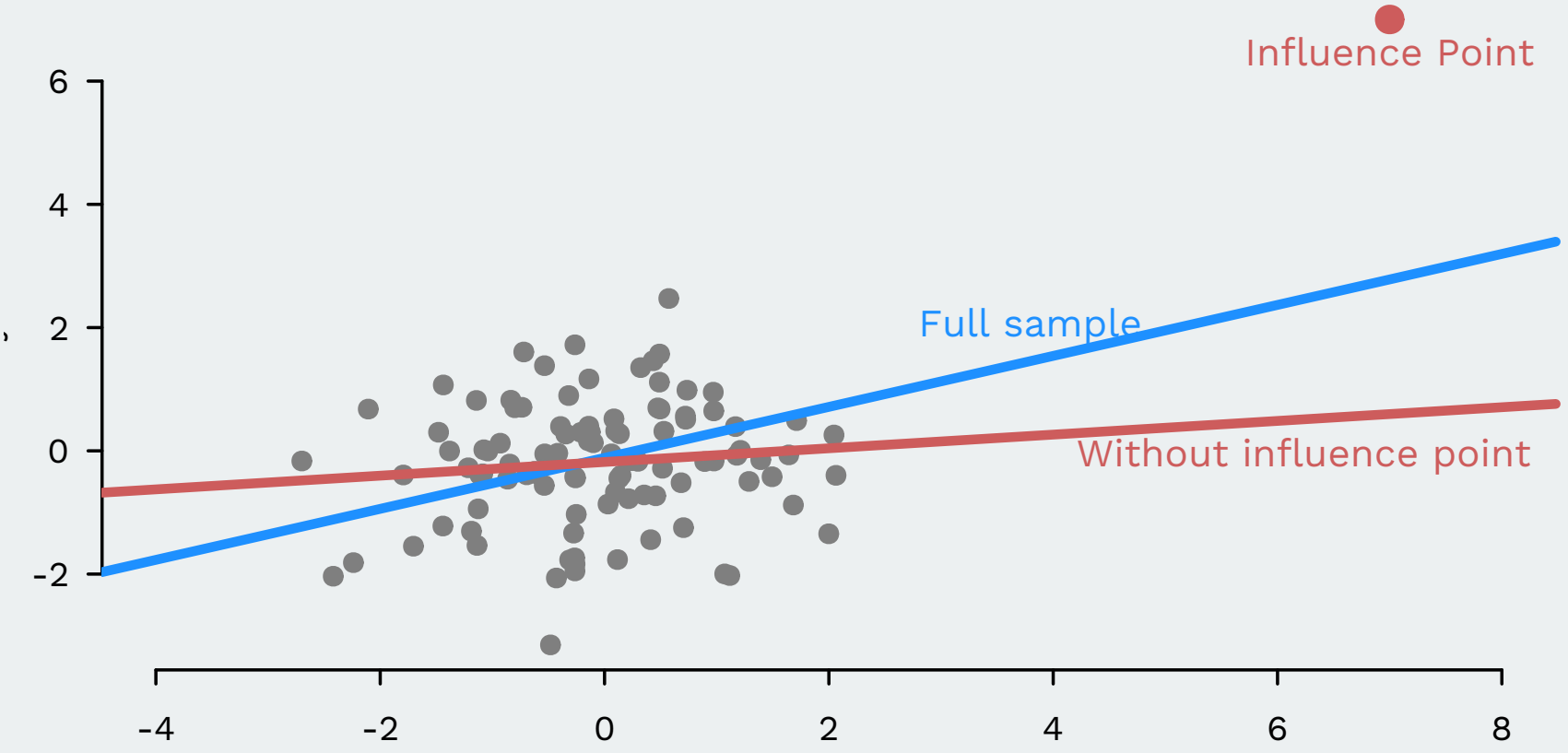
```
head(hatvalues(mod), 5)
```

```
##       1       2       3       4       5
## 0.04179 0.02285 0.22066 0.01556 0.01493
```

Buchanan hats



Influence points



- An influence point is one that is both an outlier and a leverage point.
- Extreme in both the x and y dimensions
- Causes the regression line to move toward it (bias?)

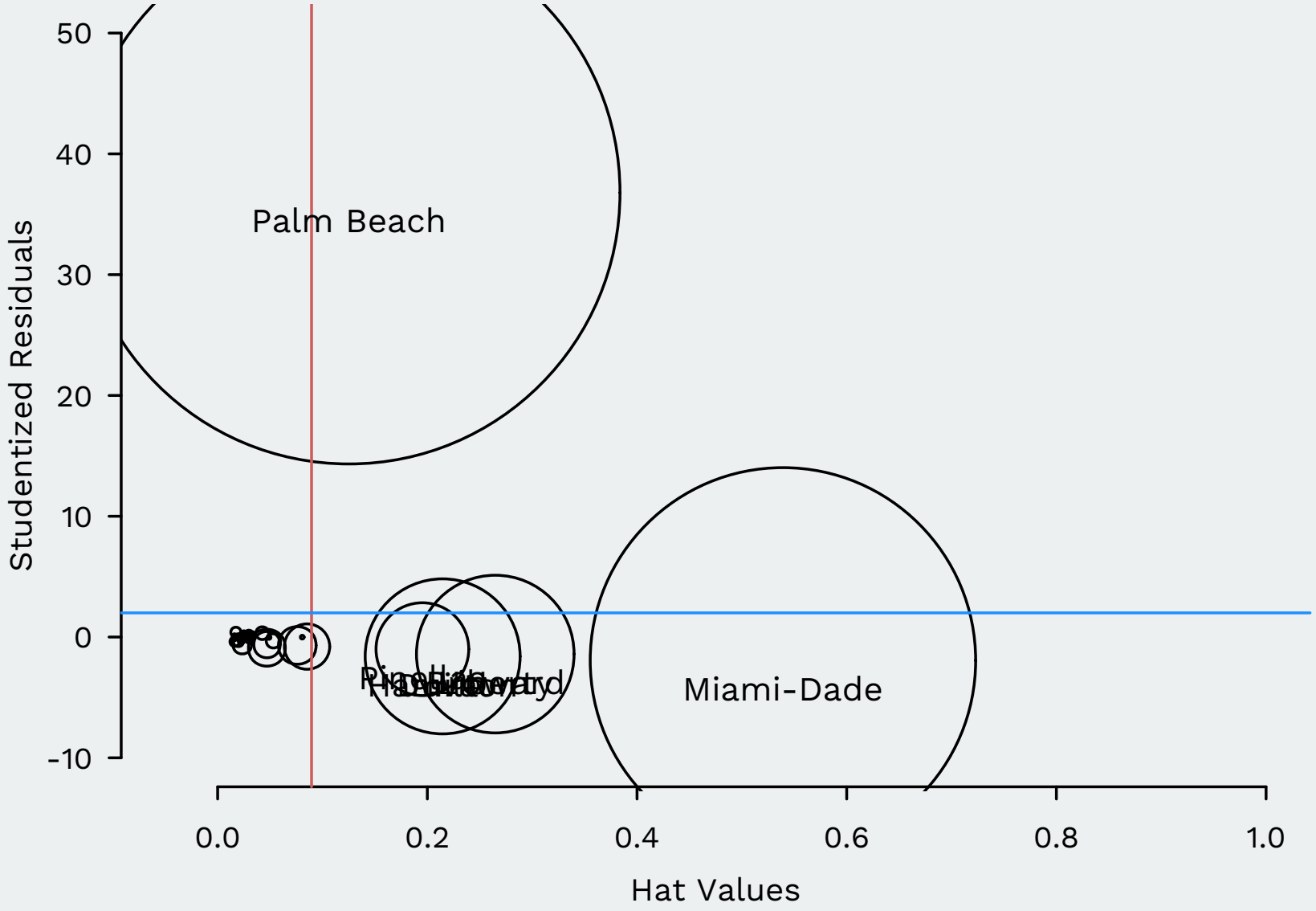
Overall measures of influence

- A measure of influence for each observation is Cook's distance:

$$D_i = \frac{\hat{u}_i'}{k+1} \times \frac{h_i}{1-h_i}$$

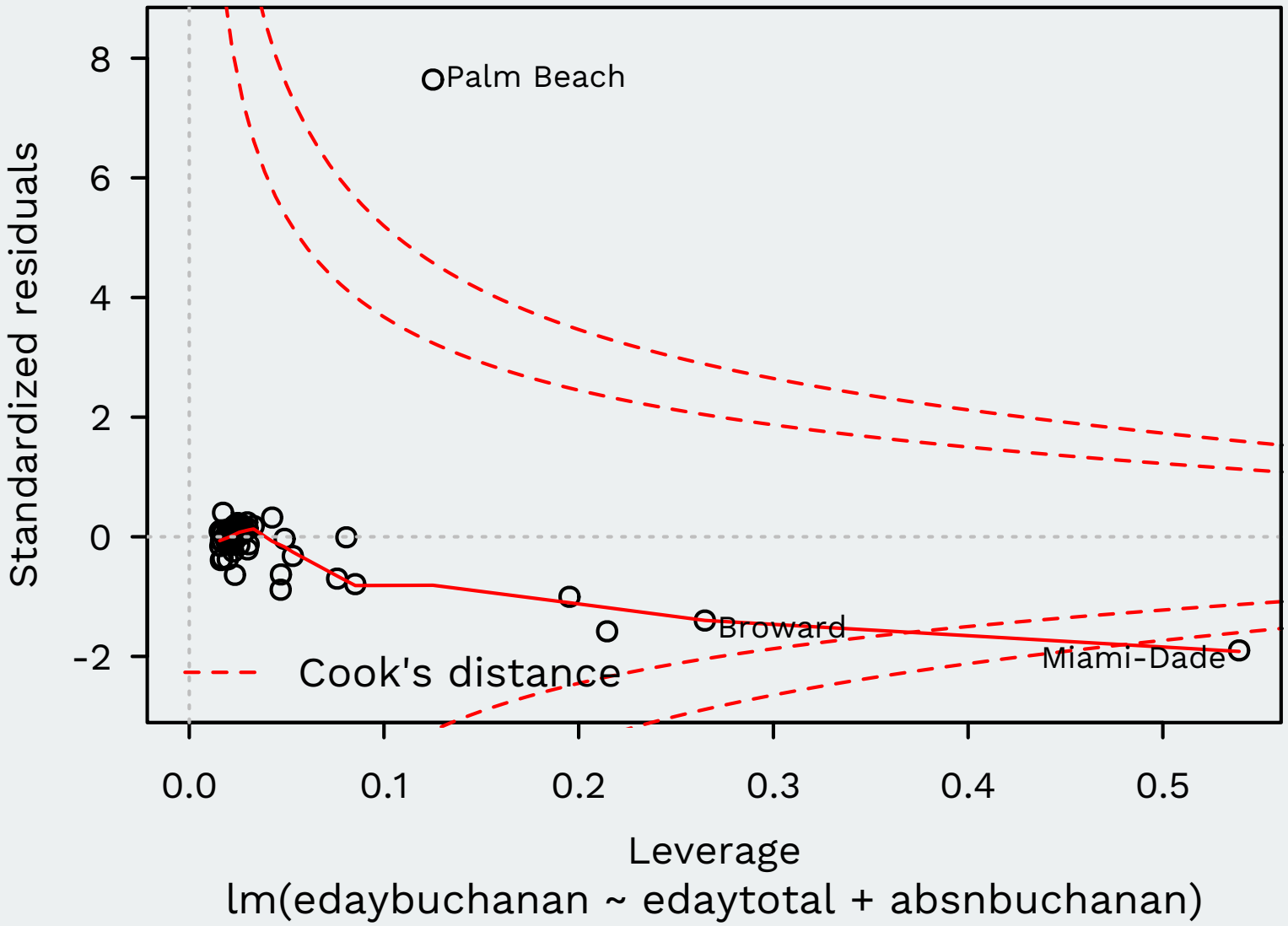
- Remember here that $\hat{u}_i'$ is the standardized residual and h_i is the hat value.
- Basically this is "outlier × leverage"
- $D_i > 4/(n-k-1)$ considered "large"
- Influence plot:
  - x-axis: hat values, h_i
  - y-axis: studentized residuals, $\hat{u}_i^*$
  - size of points: Cook's distance

Influence Plot Buchanan

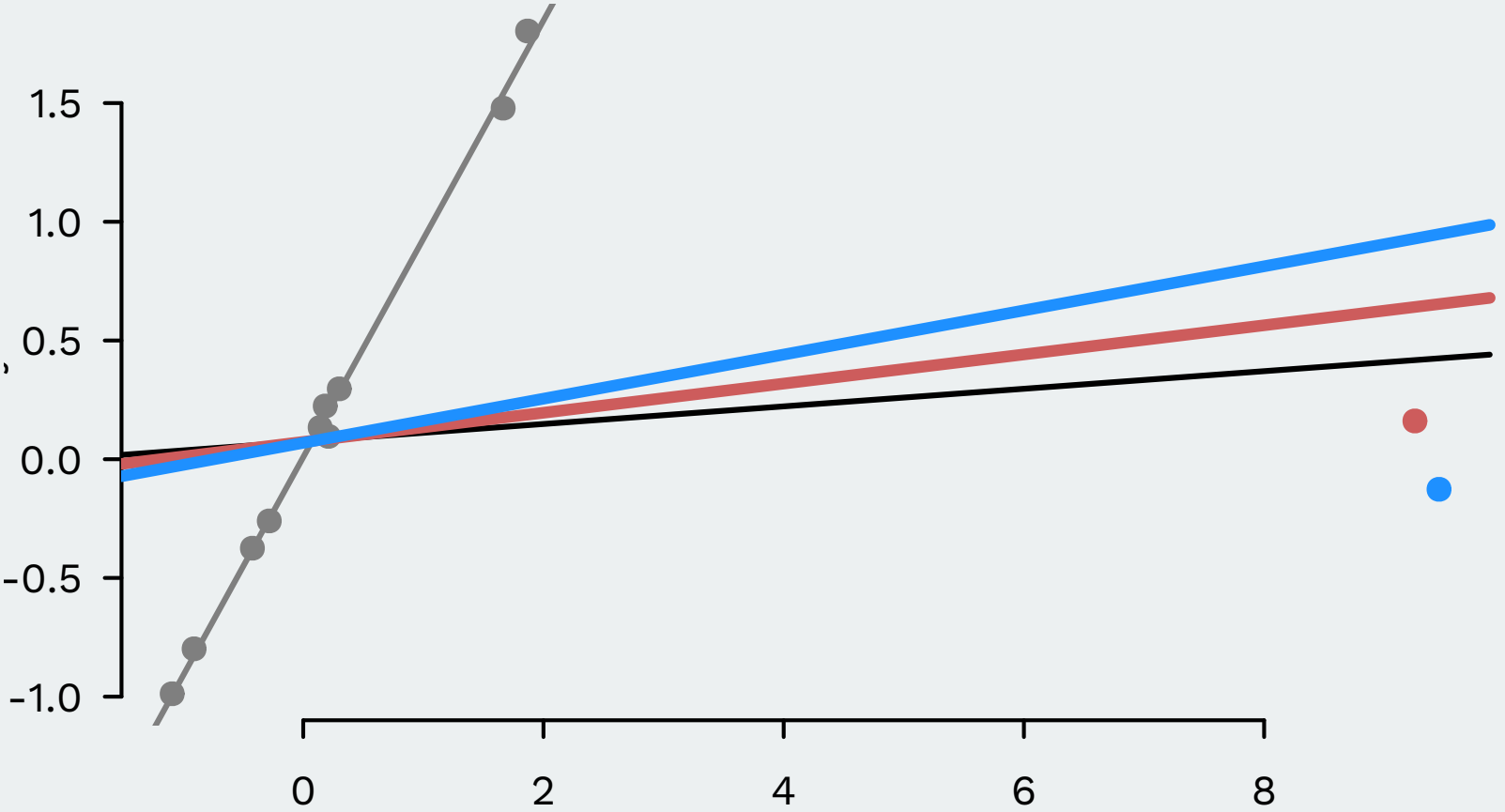


Influence plot from lm output

```
plot(mod3, which = 5, labels.id = flvote$county)
```



Limitations of the standard tools



- What happens when there are two influence points?
- Red line drops the red influence point
- Blue line drops the blue influence point
- Neither of the “leave-one-out” approaches helps recover the line

What to do about outliers and influential units?

- Is the data corrupted?
  - Fix the observation (obvious data entry errors)
  - Remove the observation
  - Be transparent either way
- Is the outlier part of the data generating process?
  - Transform the dependent variable ($\log(y)$)
  - Use a method that is robust to outliers (robust regression, least absolute deviations)

Summary

- For nonnormality, influential points, and nonlinearity:
  - Check your data! `summary()`, `plot()`, etc
  - Use transformations to make assumptions more plausible
  - Add covariates to help account for non-identical distributions
- Next week:
  - What if we have heteroskedastic data?