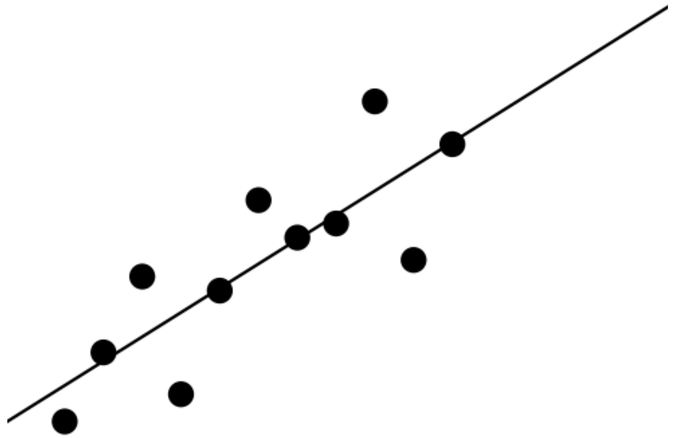
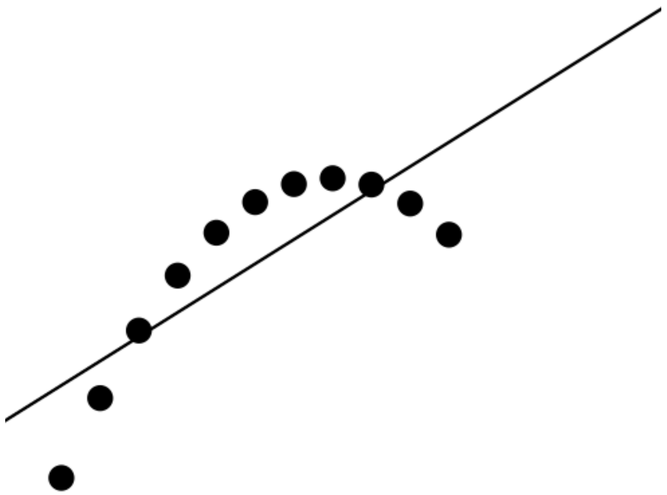
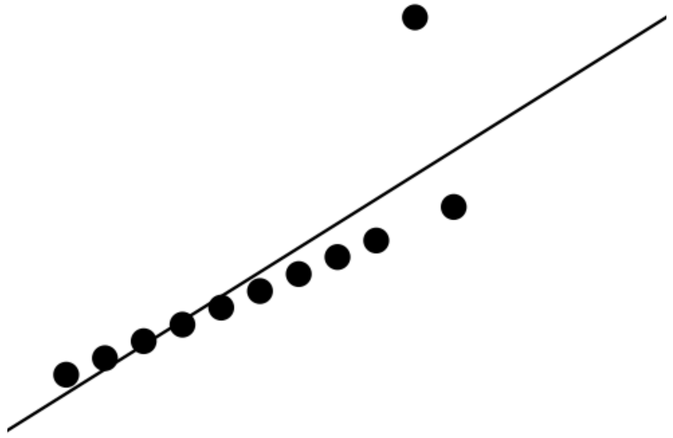
incorporating visualization into statistics education

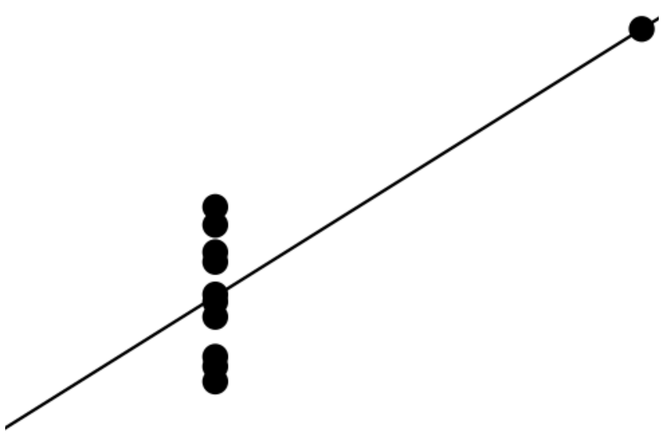




visualization forum

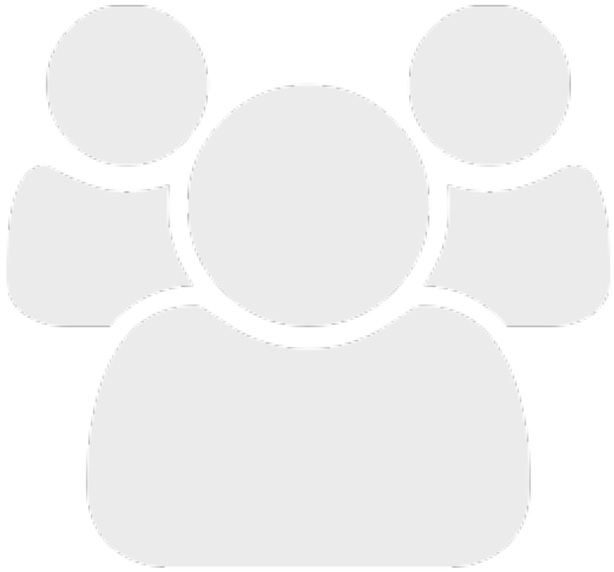
mine çetinkaya-rundel
march 4, 2016





part 1 -
teaching
visualization

part 2 -
using
visualization
to teach



sta 101 - dasi

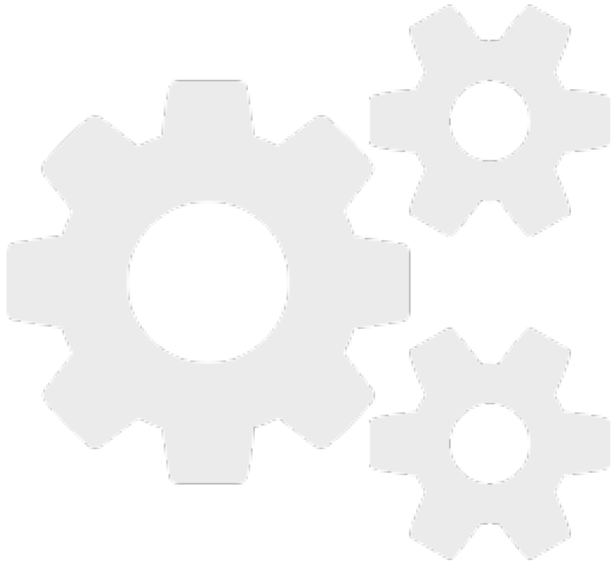
required intro stat course for (mostly) social science majors

sta 112fs - data science

first-year ugrad course in the “What if?” focus cluster

coursera - dasi

MOOC with similar content to sta 101



+

RStudio®

relevant
beyond the
intro classroom

built-in
reproducibility

free and
open-source

part 1 - teaching visualization

statistical:

visualize

`Y ~ X | Z`

with ease

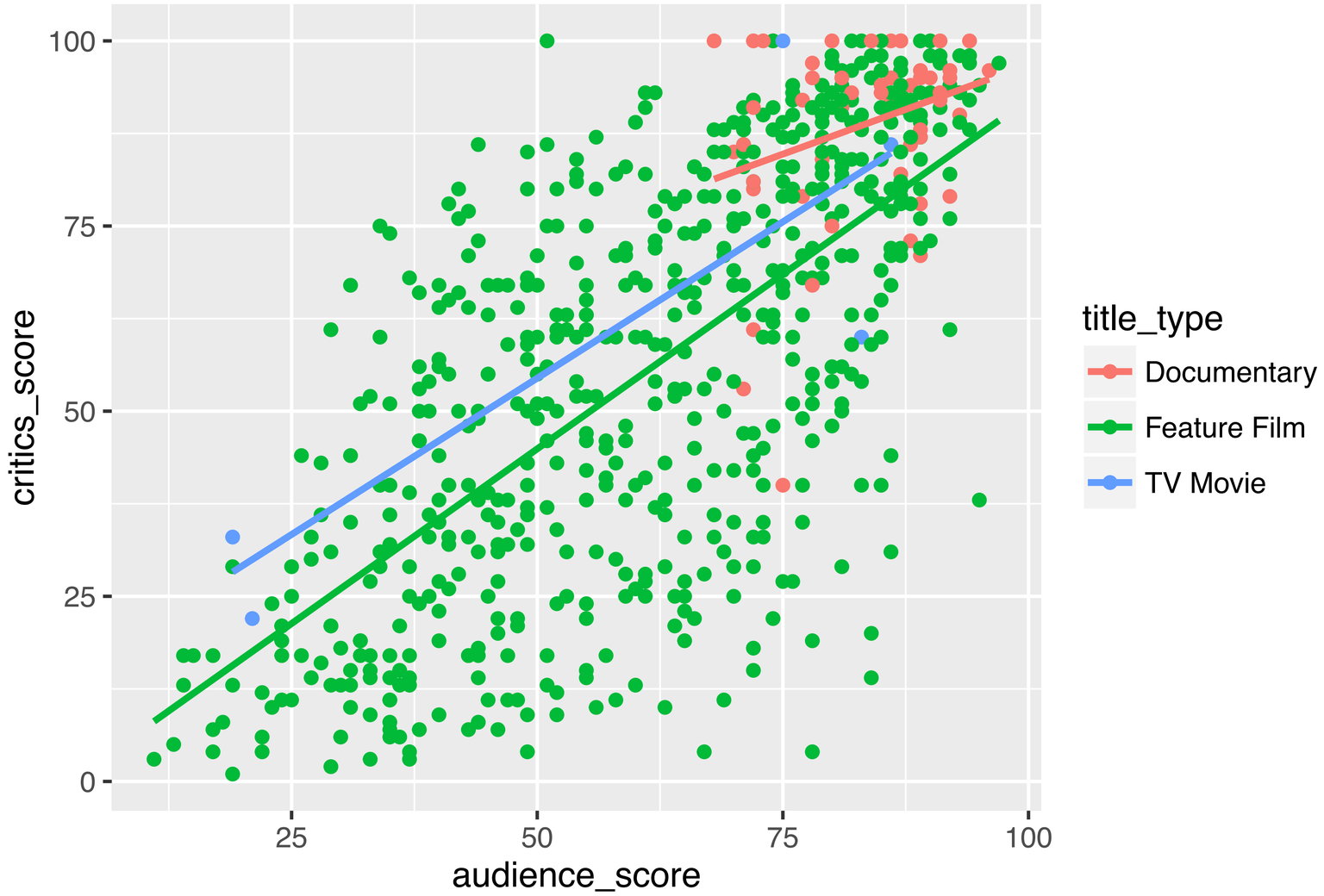
technical:

human readable syntax

bonus:

aesthetically pleasing

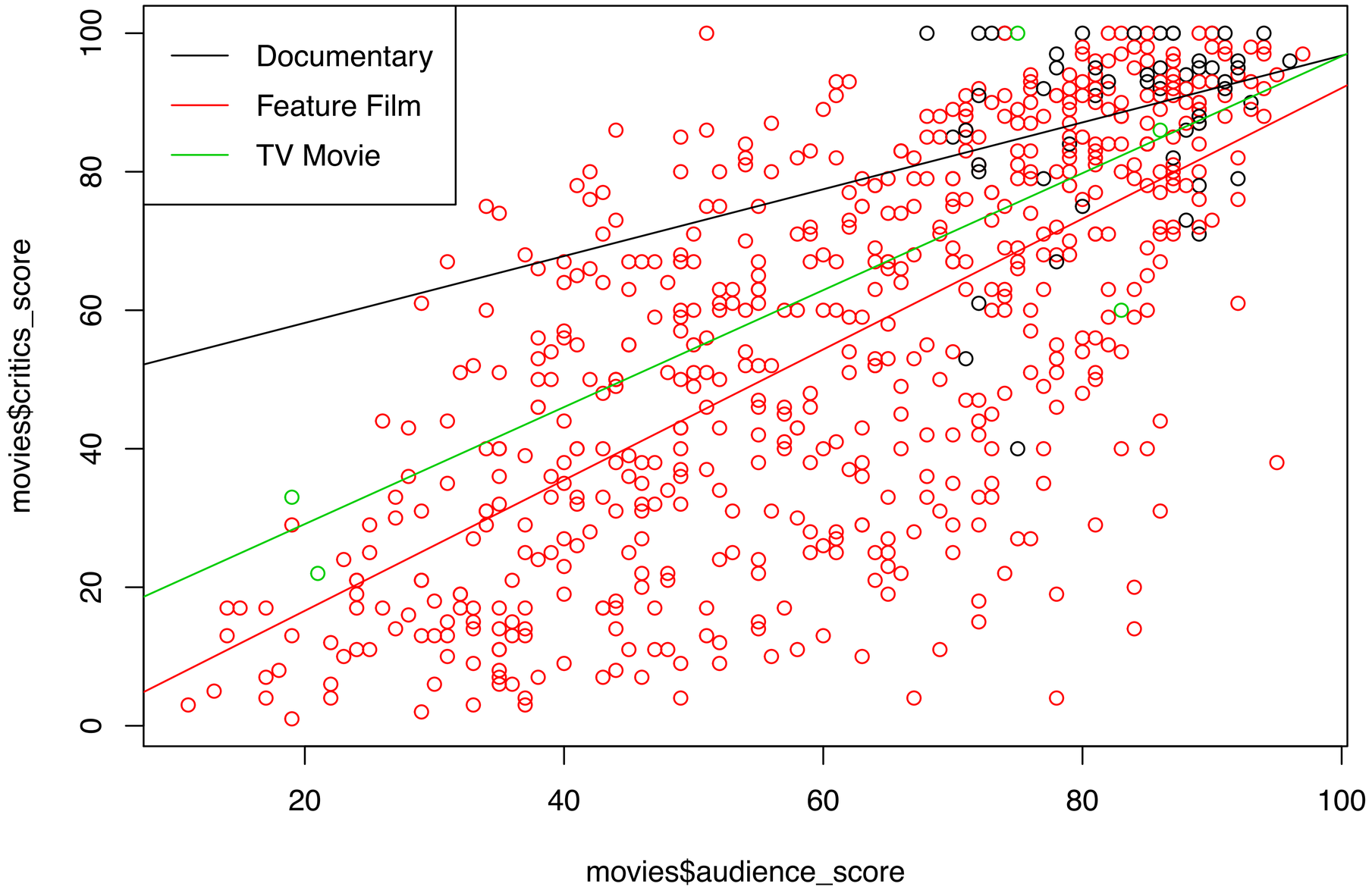
```
qplot(data = movies, x = audience_score, y = critics_score,
      color = title_type) +
   stat_smooth(method = "lm", se = FALSE)
```



ggplot2

```r
plot(y = movies$critics_score, x = movies$audience_score,
     col = movies$title_type)

doc <- movies[movies$title_type == "Documentary", ]
ff <- movies[movies$title_type == "Feature Film", ]
tv <- movies[movies$title_type == "TV Movie", ]

m_doc <- lm(critics_score ~ audience_score, data = doc)
m_ff <- lm(critics_score ~ audience_score, data = ff)
m_tv <- lm(critics_score ~ audience_score, data = tv)

abline(m_doc, col = 1)
abline(m_ff, col = 2)
abline(m_tv, col = 3)

legend("topleft", levels(d$title_type), col = c(1,2,3), lty = 1)
```

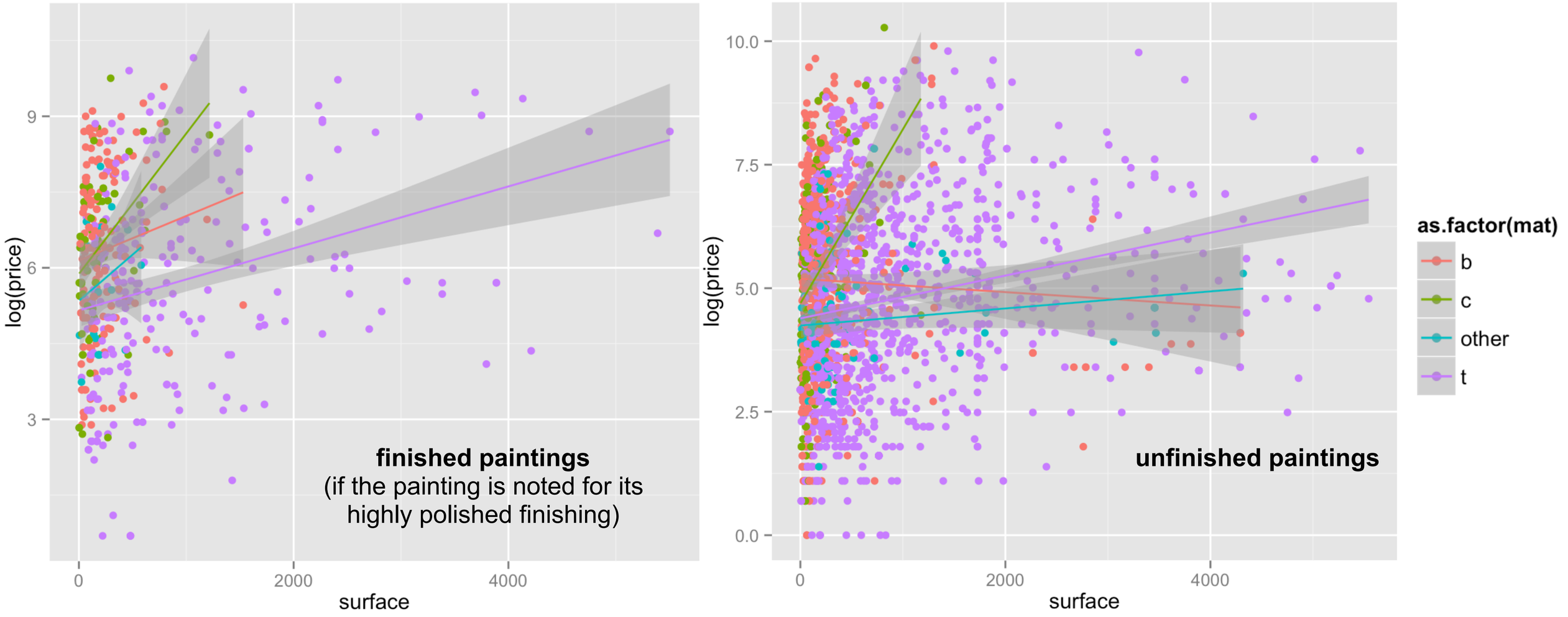
base


Documentary
Feature Film
TV Movie
movies$critics_score
movies$audience_score
0
20
40
60
80
100
20
40
60
80
100

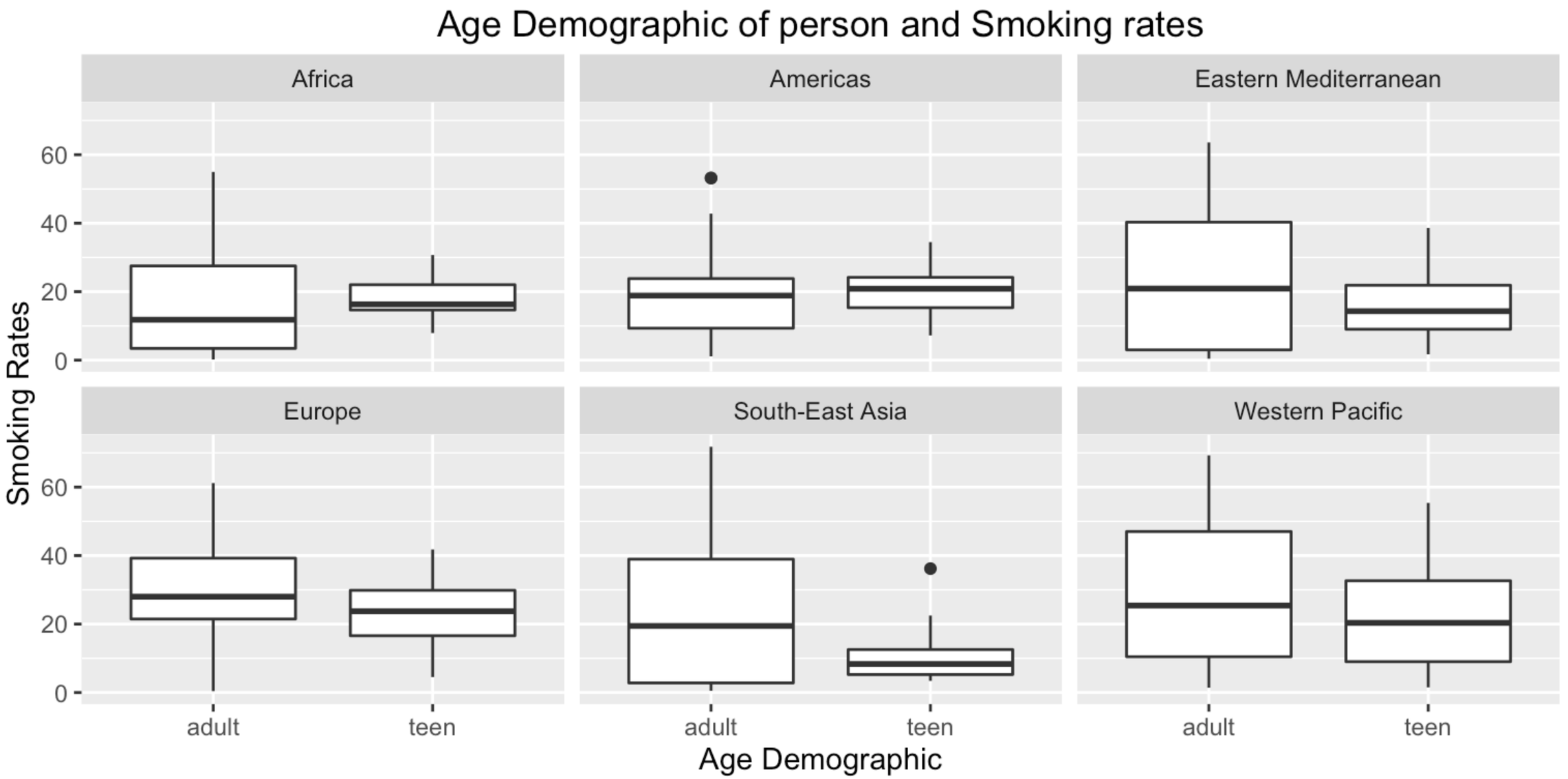
base

team: MINEstrone soup

Price vs. surface area of paintings, by material type



team: Mean Girls



interactive
visualizations
in R

R syntax

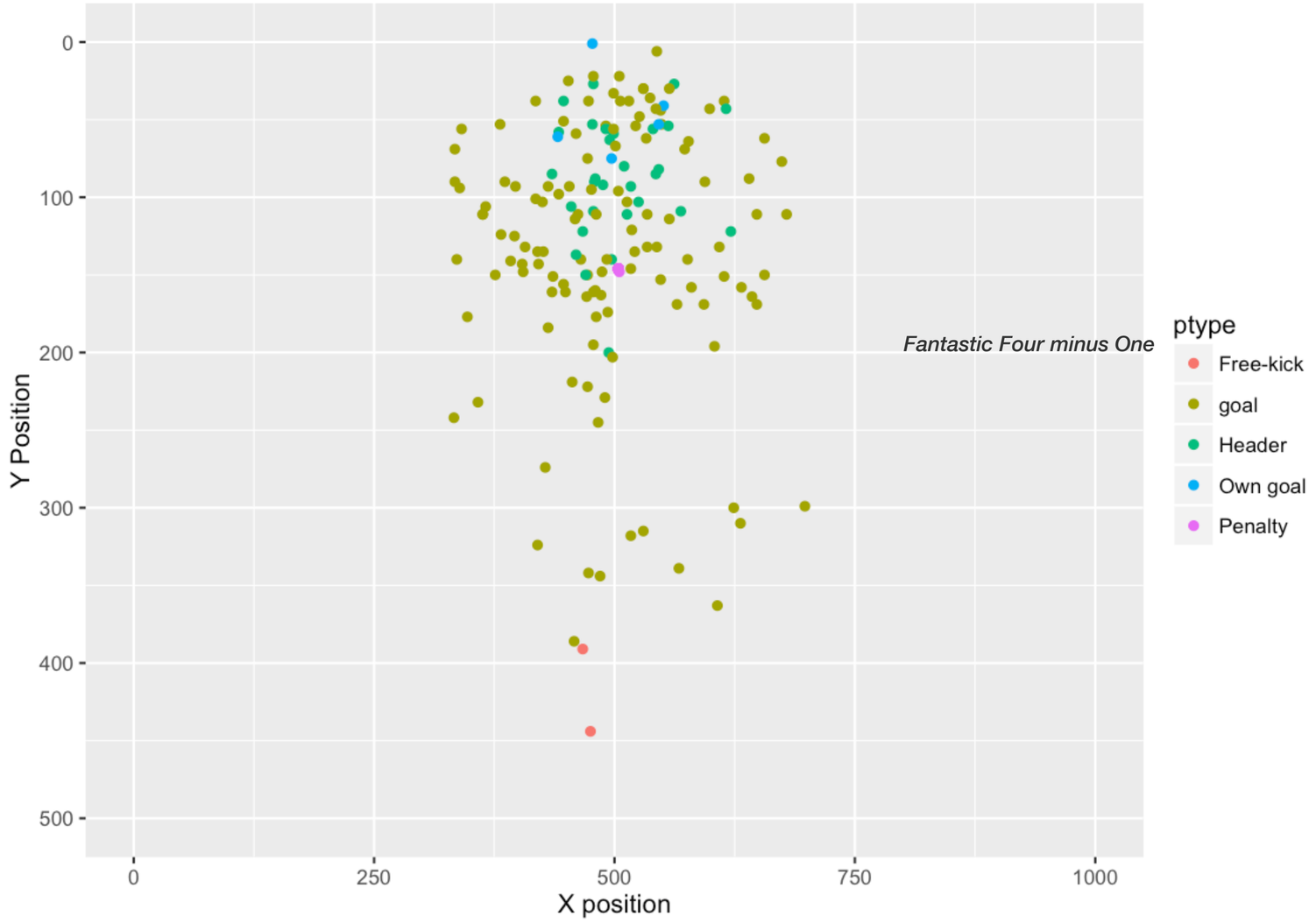
+

a bit more

access to all
statistical
funtionality in R

team: Fantastic Four minus One

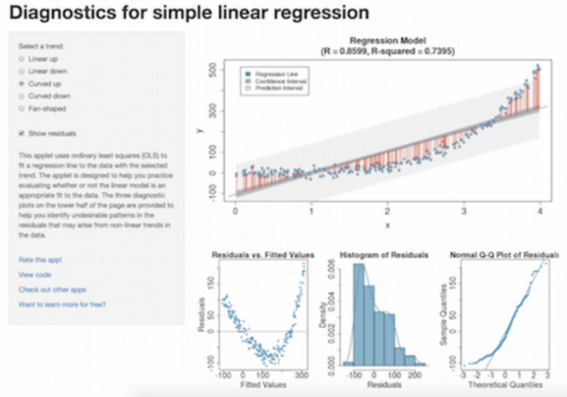
Kick locations by Type



Select the Goal Type

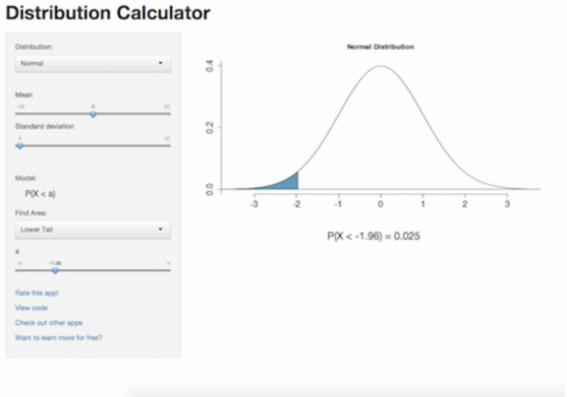
- (•) All
- ( ) Free-Kick
- ( ) Goal
- ( ) Header
- ( ) Own Goal
- ( ) Penalty

part 2 - using visualization to teach



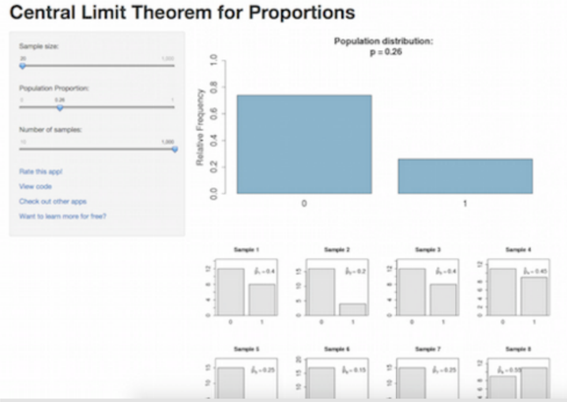
Diagnostics for simple linear regression

Diagnostic plots for simple linear regression where the variables have a linear up/down, curved up/down, and fan-shaped relationship.



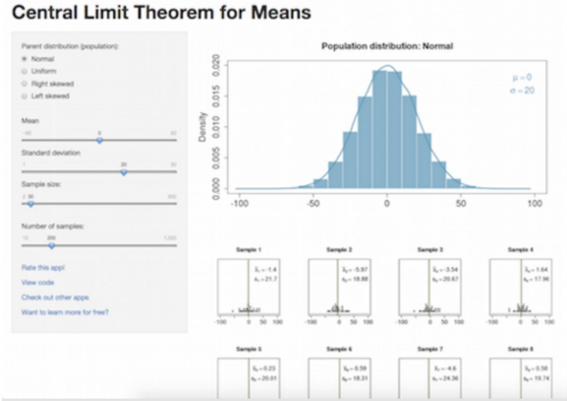
Distribution calculator

Calculate probabilities under the normal, binomial, t, F, and chi-square distributions.



CLT for proportions

Central limit theorem for distribution of sample proportions for varying sample sizes and proportions of success.



CLT for means

Central limit theorem for distributions of sample means of varying sample sizes from normal, uniform, right skewed, and left skewed populations.

http://www.stat.duke.edu/~mc301/shinyed

Central Limit Theorem for Proportions

Sample size:

2 500 1,000

2 102 202 302 402 502 602 702 802 902 1,000

Population Proportion:

0 0.2 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Number of samples:

10 1,000

10 109 208 307 406 505 604

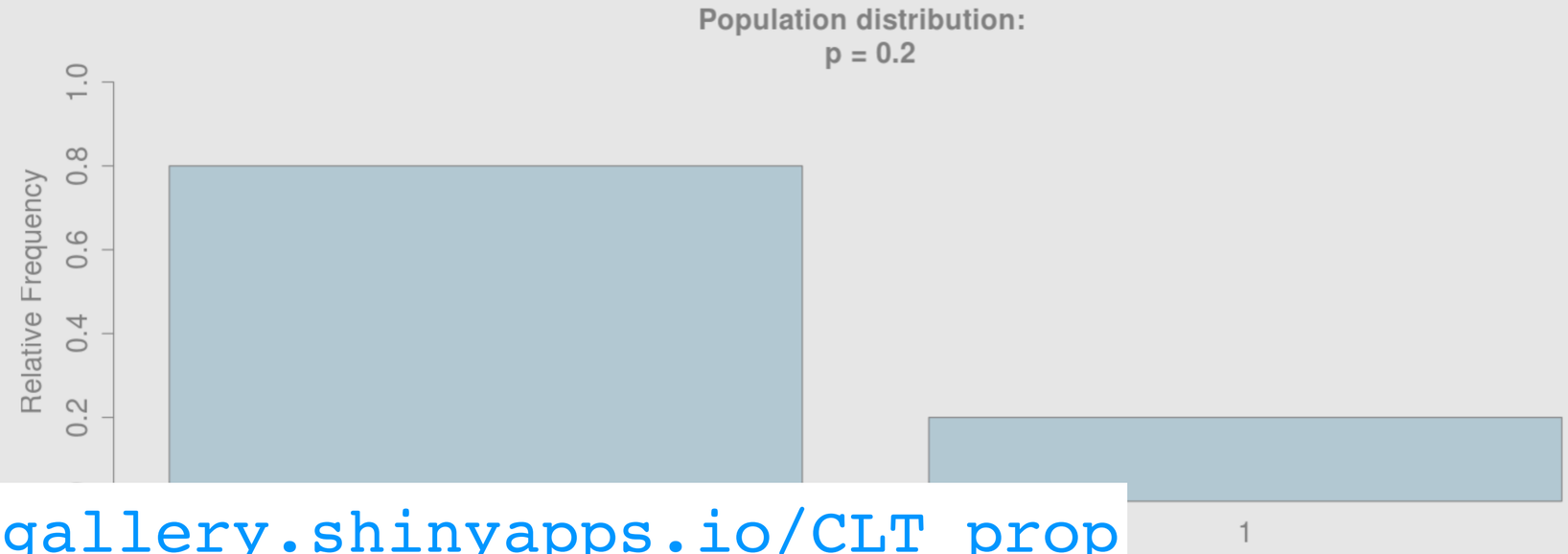
Rate this app!

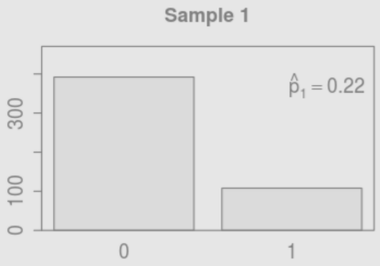
View code

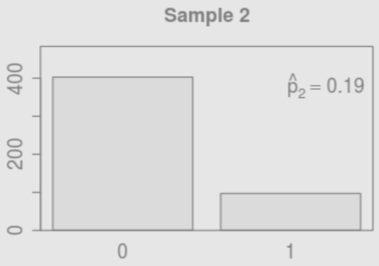
Check out other apps

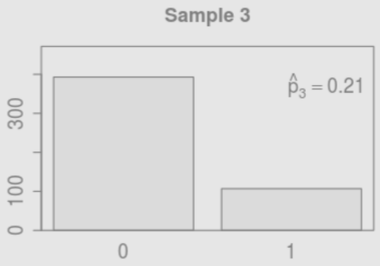
Want to learn more for free?

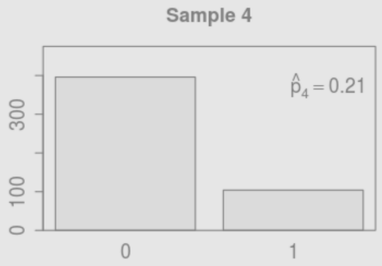
https://gallery.shinyapps.io/CLT_prop

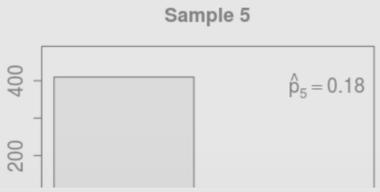


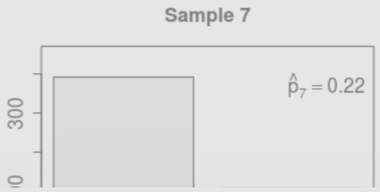












Diagnostics for simple linear regression

Select a trend:

- Linear up
- Linear down
- Curved up
- Curved down
- Fan-shaped

Show residuals

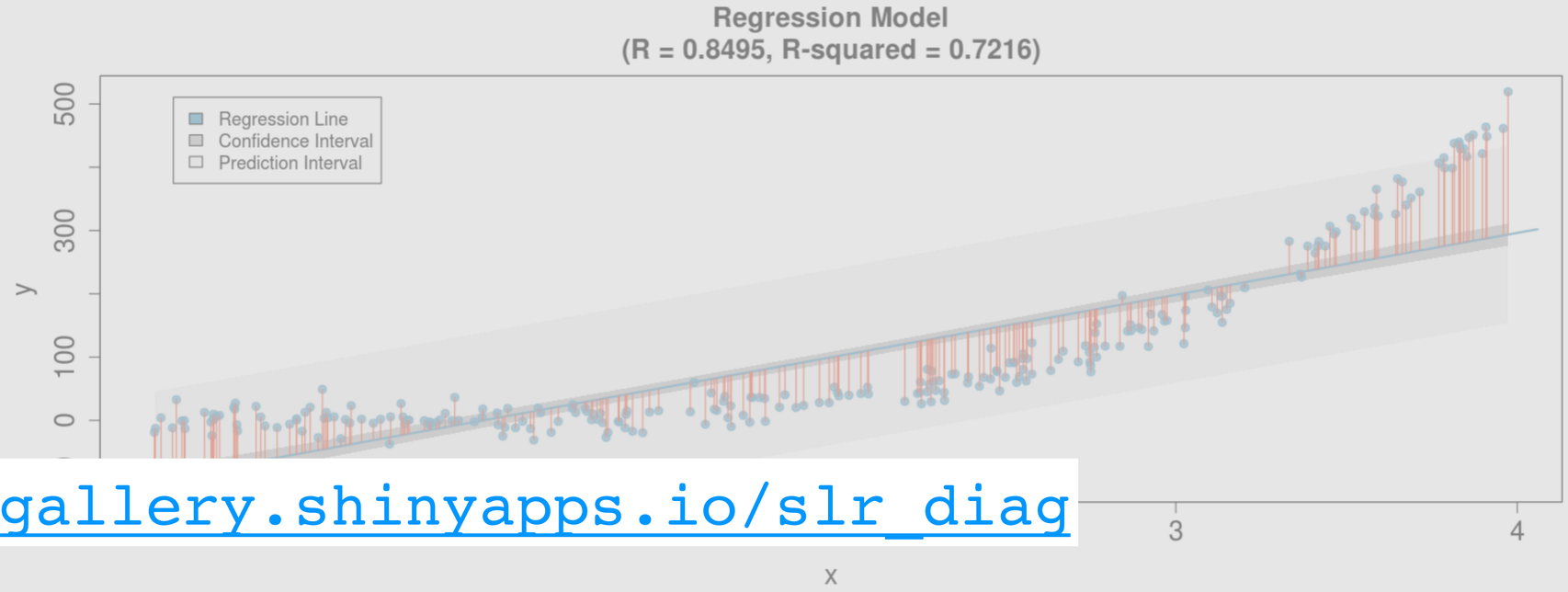
This applet uses ordinary least squares (OLS) to fit a regression line to the data with the selected trend. The applet is designed to help you practice evaluating whether or not the linear model is an a
The three diagnostic plots on the lower half of th
help you identify undesirable patterns in the resid
non-linear trends in the data.

Rate this app!

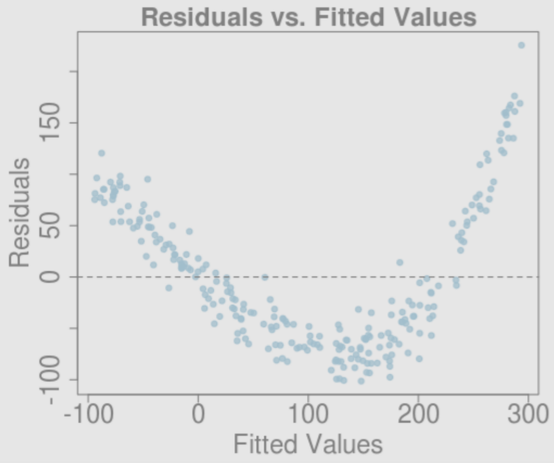
View code

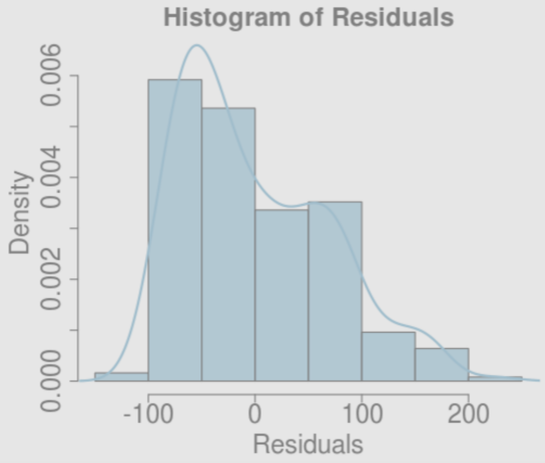
Check out other apps

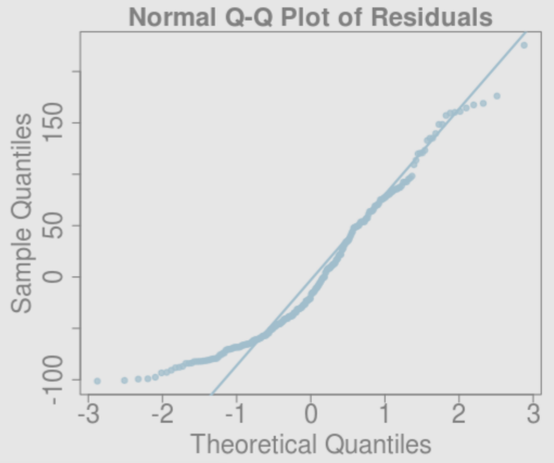
Want to learn more for free?



https://gallery.shinyapps.io/slr_diag







Google Chrome

thank you!