# Using the Illinois Report Card Data to Teach Statistics

MMC Conference of Workshops

*Frank Briody*
*Prospect High School*
*frankbriody@gmail.com*

*2/1/2020*

## Contents

# 1 Variables

The ISBE raw data file *rx17.txt* contains 1,471 variables. The variable definitions are in the Excel file *RC17_layout.xlsx* and have been categorized into the groups shown below. The first number represents available variables in each group while the second is the number actually imported into the processed data file. The import script produces 316 variables from 20 of the 21 categories for all 3,796 Illinois public schools. (None of the NAEP variables were imported.) Usable files will be discussed in section 4.2.1.

---

School information (13 variables;12 imported)
Student demographics (396;71)
ACT (44;11)
Instructional setting (92;2)
Teacher and admin statistics (78;26)
District financial (67;40)
Region and legislative (3;2)
National Assmnt. of Educ. Progress (NAEP) (184;0)
College and Career readiness (16;3) CTE (4;1)
Advanced coursework (12;3)
AP courses (168;42)
IB courses (168;42)
Dual credit (168;42)
AP exams (36;12)
Post secondary remediation (4;1)
Response rate (5E survey) (4;2)
Health and wellness (3;1)
Teacher Attendance (4;1)
Teacher Evaluation (2;1)
School District Count (3;1)

---

# 2 Descriptive Statistics via State Demographics

## 2.1 Categorical Count (Raw)

```
school_type <- rc17 %>%
  count(SCHOOL_TYPE_NAME, sort = TRUE) %>%
  mutate(rel_freq = n/sum(n))
school_type
```

```
## # A tibble: 4 x 3
##   SCHOOL_TYPE_NAME     n rel_freq
##   <chr>            <int>    <dbl>
## 1 ELEMENTARY        2406   0.634
## 2 HIGH SCHOOL        644   0.170
## 3 MIDDLE SCHL        604   0.159
## 4 CHARTER SCH        142   0.0374
```

## 2.2 Categorical Count (Formatted)

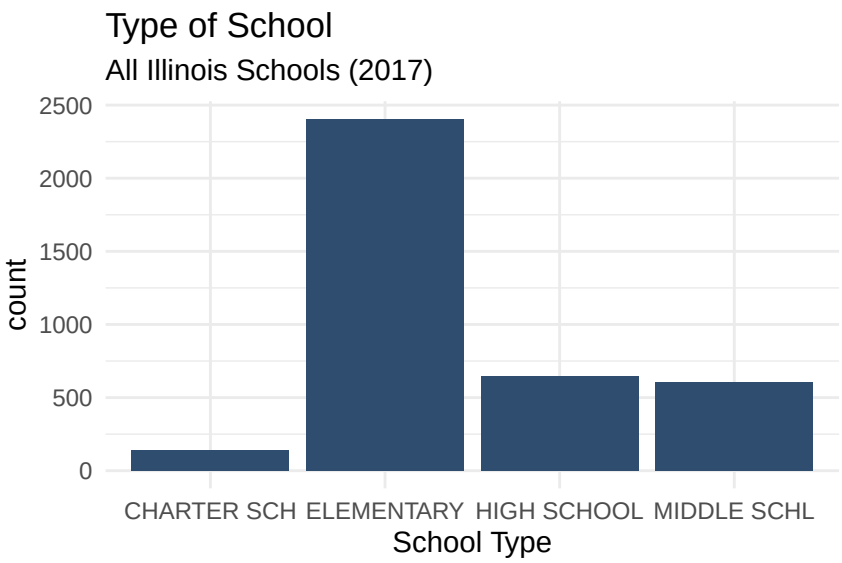
```
kable(school_type) %>%
  kable_styling(bootstrap_options = "striped", full_width = F)
```

| SCHOOL_TYPE_NAME | n | rel_freq |
|---|---|---|
| ELEMENTARY | 2406 | 0.6338251 |
| HIGH SCHOOL | 644 | 0.1696523 |
| MIDDLE SCHL | 604 | 0.1591149 |
| CHARTER SCH | 142 | 0.0374078 |

## 2.3 Categorical Plot

```
ggplot(rc17, aes(x=factor(SCHOOL_TYPE_NAME)))+
  geom_bar(fill="#2F4E6F")+
```

```
  labs(title = "Type of School", x = "School Type", subtitle = "All Illinois Schools (2017)") +
  theme_minimal()
```

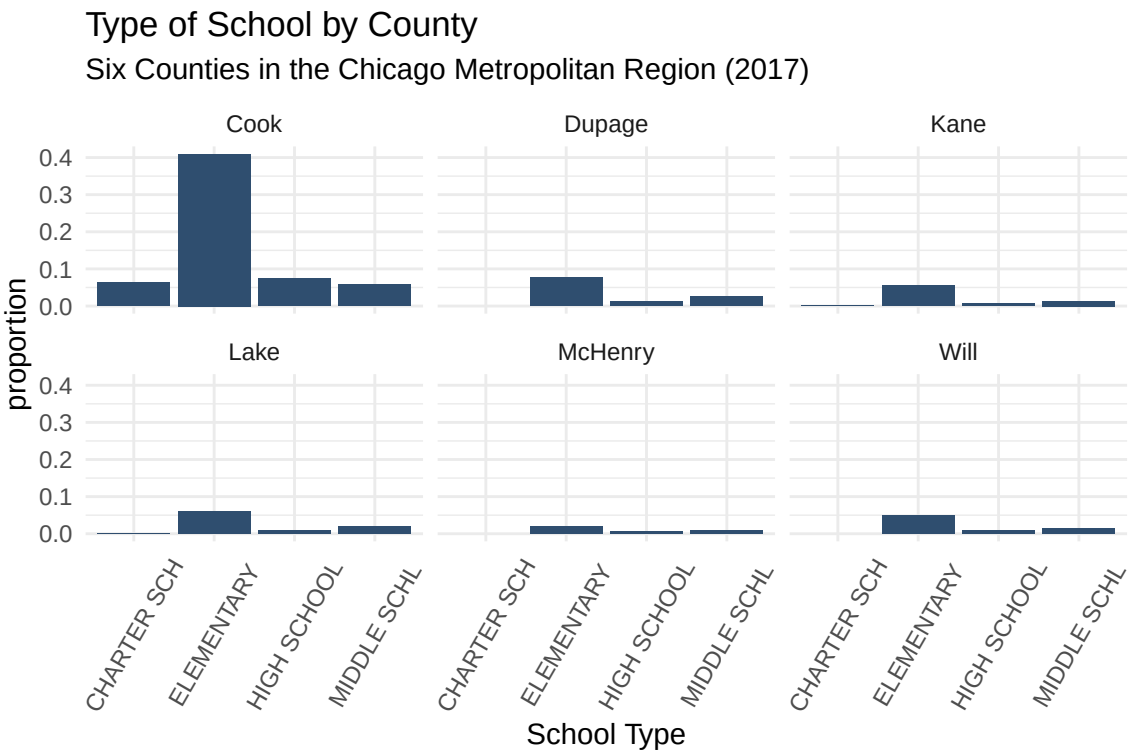



## 2.4 Categorical Analysis I

Write a short analysis for the types of schools in the state of Illinois.

## 2.5 Categorical Analysis II

```
rc17 %>%
filter(COUNTY == "Dupage" | COUNTY == "Will" | COUNTY == "Kane" |
         COUNTY == "Lake" | COUNTY == "Cook"| COUNTY == "McHenry") %>%
ggplot(aes(x=factor(SCHOOL_TYPE_NAME), y = (..count..)/sum(..count..))) +
  geom_bar(fill="#2F4E6F")+
  facet_wrap(~COUNTY, nrow = 2) +
  labs(title = "Type of School by County",
       x = "School Type",
       y = "proportion",
       subtitle = "Six Counties in the Chicago Metropolitan Region (2017)") +
  theme_minimal()  +
  theme(axis.text.x = element_text(angle = 60, vjust = 0.5))
```




Write a short analysis for the types of schools in the six county region.

# 3 Software

## 3.1 R and RStudio

R is open source, free, industry-standard statistical analysis software that was first introduced in 1993. (R is an adaptation of the S languge that was invented at Bell Labs in 1976.) RStudio, which arrived in 2011, is a free development environment for using R. Both R and RStudio have a bit of an initial learing curve, but mastering a few basic commands opens up a world of analysis options. The RStudio installation page [https://rstudio.com/products/rstudio/download/#download] provides instructions for setting up both R and RStudio. Suggestions for learning R are in the Appendix.

## 3.2 Fathom

Key Curriculum Press, the creators of The Geometer's Sketchpad software, sells Fathom ($39 USD) at [https://fathom.concord.org]. Fathom is graphical statistical analysis software which accomplishes most tasks through a drag and drop interface. Its primary audience is teachers and students, but has reached end-of-life status. It can still be purchased and used but no further developments are expected.

## 3.3 Others

Excel, JMP, and SAS are other software platforms that can be used for analysis but will not be discussed here. The report card data will be provided in .csv format which can be imported by these programs.

# 4 Data Import

## 4.1 Data Files

INSERT GRAPHIC HERE

### 4.1.1 Super Easy Method

All graphs from this presentation are available for download at [INSERT MMC GITHUB HTML LINK]. ???Look in the appendix section.

### 4.1.2 Easy Method

A reasonably sized (2.7MB) data file containing 316 variables for the 2,049 schools in the Chicagoland six county region is avaialble at [GITHUB LINK]. The six counties are Cook, Dupage, Kane, Lake, McHenry and Will. This data file can be imported into RStudio, Fathom, Excel, etc.

### 4.1.3 Some Variables for All Schools

I used the import script `import_rc17.txt` to select 316 variables for all 3,796 Illinois schools. A 5MB data file is available at []. Again, this data file can be imported into RStudio, Fathom, Excel, etc.

### 4.1.4 Starting from Scratch

The original raw data is available on the ISBE Report Card Data Library web page [https://www.isbe.net/Pages/Illinois-State-Report-Card-Data.aspx]. You will need to download both the fixed width data file (rc17.txt 2.4MB becomes 35.4MB) and the variable definitions (RC17_layout.xlsx). Use and/or modify import_rc17.txt [https://github.com/fbriody/MMC2020] to get a subset of the data into RStudio.

Subsetting and exporting is a two step process. First, use R to create the subset:

```
sixco <- rc17 %>%
  filter(COUNTY %in% c("Cook", "Lake", "Will", "Kane", "McHenry", "Dupage"))
```

Then, export:

```
write.csv(sixco,"sixco.csv", row.names = FALSE)
```

The resulting .csv can be imported by another software platform. (Note there is no need to export if you are staying in Studio. Just refer to you r new subset, in this case `sixco`.) You can customize the above command(s) to suit your needs.

## 4.2 Processed into Fathom

### 4.2.1 Import

csv import attribute names

### 4.2.2 Displays

missing values and shift or option inspecting adding a third variable act v LowInc with school type

## 4.3 Importing .csv into Excel

Use Excel to import rc17.csv. *****CHANGE NAME to rc17_fb.csv??*****

# 5 Numeric Summaries (REORDER THESE SECTIONS)

## 5.1 Lists

### 5.1.1 Number of High Schools in the Six County Region

```
sixco %>%
  filter(SCHOOL_TYPE_NAME == "HIGH SCHOOL") %>%
  group_by(COUNTY) %>%
  summarise(count = n())
```

```
## # A tibble: 6 x 2
##   COUNTY  count
##   <chr>   <int>
## 1 Cook      151
## 2 Dupage     23
## 3 Kane       16
## 4 Lake       21
## 5 McHenry    14
## 6 Will       17
```

### 5.1.2 Single List of Scores

Subsetting based on a criteria produces a single list of scores.

```
mchenry_act <- rc17 %>%
  filter(COUNTY == "McHenry", is.na(ACT_COMP_SCHOOL) == FALSE )
mchenry_act$ACT_COMP_SCHOOL
```

```
##  [1] 22.4 19.7 18.1 23.1 22.6 23.8 22.7 24.0 21.1 19.9 22.9 22.9 21.4 21.2
```

Adding `sort()` orders the scores. Remove the comma and the `decreasing` option to priduce an increasing list.

```
sort(mchenry_act$ACT_COMP_SCHOOL, decreasing = TRUE)
```

```
##  [1] 24.0 23.8 23.1 22.9 22.9 22.7 22.6 22.4 21.4 21.2 21.1 19.9 19.7 18.1
```

### 5.1.3 Lake County ACT Scores (Ordered and Formatted)

```
lake_ACT <- rc17 %>%
  filter(SCHOOL_TYPE_NAME == "HIGH SCHOOL", COUNTY == "Lake") %>%
  arrange(desc(ACT_COMP_SCHOOL)) %>%
  select(COUNTY, SCHOOL_NAME, ACT = ACT_COMP_SCHOOL)
kable(lake_ACT)
```

Table 1: Types of School Districts

| | Dupage | Lake |
|---|---|---|
| LARGE | 182 | 150 |
| MEDIUM | 52 | 38 |
| SMALL | 0 | 4 |

| COUNTY | SCHOOL_NAME | ACT |
|---|---|---|
| Lake | Adlai E Stevenson High School | 26.9 |
| Lake | Deerfield High School | 26.4 |
| Lake | Lake Forest High School | 26.3 |
| Lake | Libertyville High School | 25.9 |
| Lake | Highland Park High School | 25.2 |
| Lake | Vernon Hills High School | 25.1 |
| Lake | Lake Zurich High School | 24.9 |
| Lake | Barrington High School | 24.8 |
| Lake | Grayslake Central High School | 23.3 |
| Lake | Lakes Community High School | 22.6 |
| Lake | Grayslake North High School | 22.4 |
| Lake | Warren Township High School | 22.1 |
| Lake | Wauconda High School | 21.8 |
| Lake | Antioch Comm High School | 21.7 |
| Lake | Mundelein Cons High School | 21.4 |
| Lake | Grant Community High School | 21.3 |
| Lake | New Tech High - Zion-Benton East | 20.1 |
| Lake | Zion-Benton Twnshp Hi Sch | 18.6 |
| Lake | Waukegan High School | 17.9 |
| Lake | Round Lake Senior High School | 17.8 |
| Lake | North Chicago Community High Sch | 17.5 |

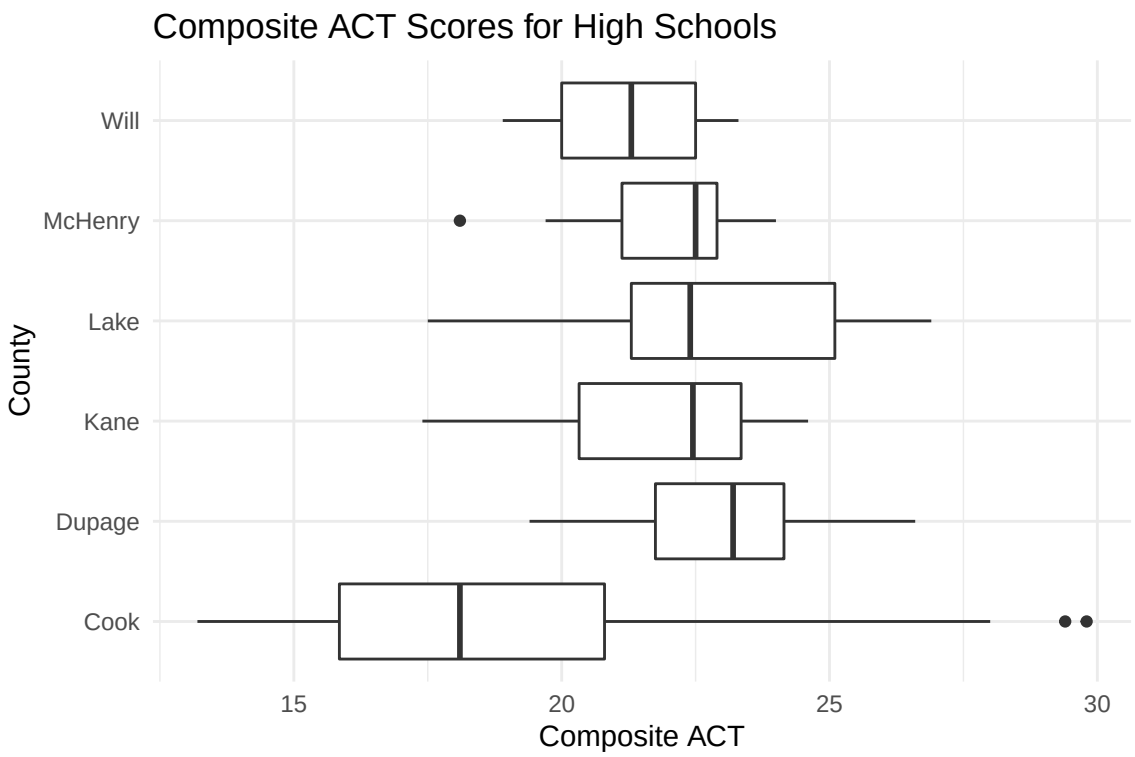
Create a boxplot for Lake County ACT scores. How could you compare to DuPage county?

## 5.2 Two-Way Tables

```
district_type <- rc17 %>%
  filter(COUNTY == "Lake" | COUNTY == "Dupage") %>%
  group_by(COUNTY)

two_way <- with(district_type, table(DISTRICT_SIZE_NAME, COUNTY))

kable(two_way, caption = "Types of School Districts") %>%
  kable_styling(bootstrap_options = "striped", full_width = F)
```

## 5.3 Resistant Measures

```
## Warning: Removed 4 rows containing non-finite values (stat_boxplot).
```




## 5.4 Mean vs Median

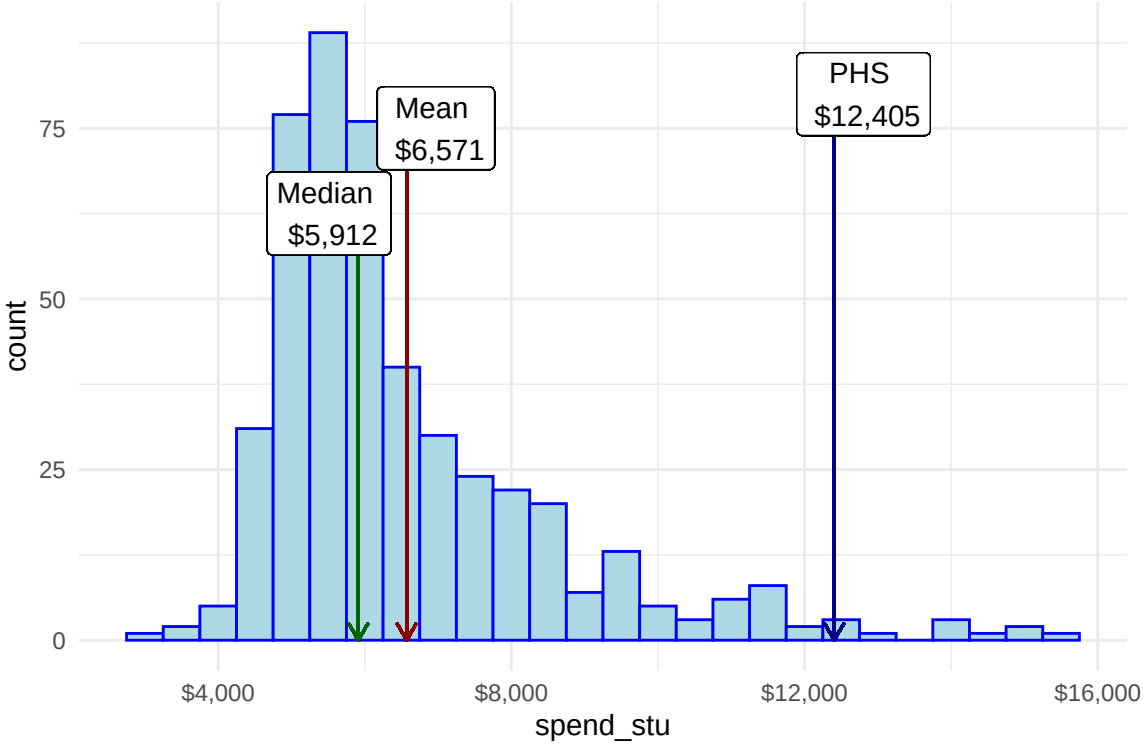
### 5.4.1 Instructional Spending per Pupil by District

#### 5.4.1.1 Numeric Summary

```
rc17 %>%
  filter(SCHOOL_TYPE_NAME == "HIGH SCHOOL") %>%
  group_by(DISTRICT_NAME) %>%
  summarise(spend_stu = mean(INSTRUCT_EXPEND_PER_PUPIL_DISTRICT201516, na.rm = TRUE)) %>%
  summary()
```

```
##  DISTRICT_NAME        spend_stu
##  Length:473         Min.   : 2975
##  Class :character   1st Qu.: 5263
##  Mode  :character   Median : 5912
##                     Mean   : 6571
##                     3rd Qu.: 7315
##                     Max.   :15535
##                     NA's   :1
```

#### 5.4.1.2 Plot




# 6 Single Values

## 6.1 Finding a School

```
rc17 %>%
  filter(str_detect(SCHOOL_NAME, "Morton")) %>%
  select(SCHOOL_ID, SCHOOL_NAME, ACT_COMP_SCHOOL)
```

```
## # A tibble: 7 x 3
##   SCHOOL_ID       SCHOOL_NAME                       ACT_COMP_SCHOOL
##   <chr>           <chr>                                       <dbl>
## 1 060162010170001 J Sterling Morton East High Sch              18.4
## 2 060162010170002 J Sterling Morton West High Sch              18.7
## 3 060162010170003 J Sterling Morton Freshman Cntr              NA
## 4 070161450022004 Morton Gingerwood Elem School                NA
## 5 150162990252844 Morton Elem Career Academy                   NA
## 6 530907090260006 Morton High School                           23.3
## 7 530907090261005 Morton Jr High School                        NA
```

## 6.2 Using a Filter

```
prospect <- rc17 %>%
  filter(str_detect(SCHOOL_NAME, "Prospect High School"))
prospect_act <- prospect$ACT_COMP_SCHOOL
prospect_act
```

```
## [1] 25
```

## 6.3 Using a Function

```
phs_value <- function(unk) {
  x <- rc17 %>%
```

```
    filter(SCHOOL_ID == "050162140170005")
  x[unk]
}

phs_value("ACT_COMP_SCHOOL")
```

```
## # A tibble: 1 x 1
##   ACT_COMP_SCHOOL
##             <dbl>
## 1              25
```

### 6.4 Single Values in Fathom

# 7 Correlation and Regression

## 7.1 Guess the Correlation

- ACT Composite vs Chronically Truant (#) Guess:_____ Actual:_____
- ACT Composite vs Chronically Truant (%) Guess:_____ Actual:_____
- ACT Composite vs Student Mobility Guess:_____ Actual:_____
- ACT Composite vs Attendance rate (%) Guess:_____ Actual:_____

## 7.2 Predicting ACT Scores

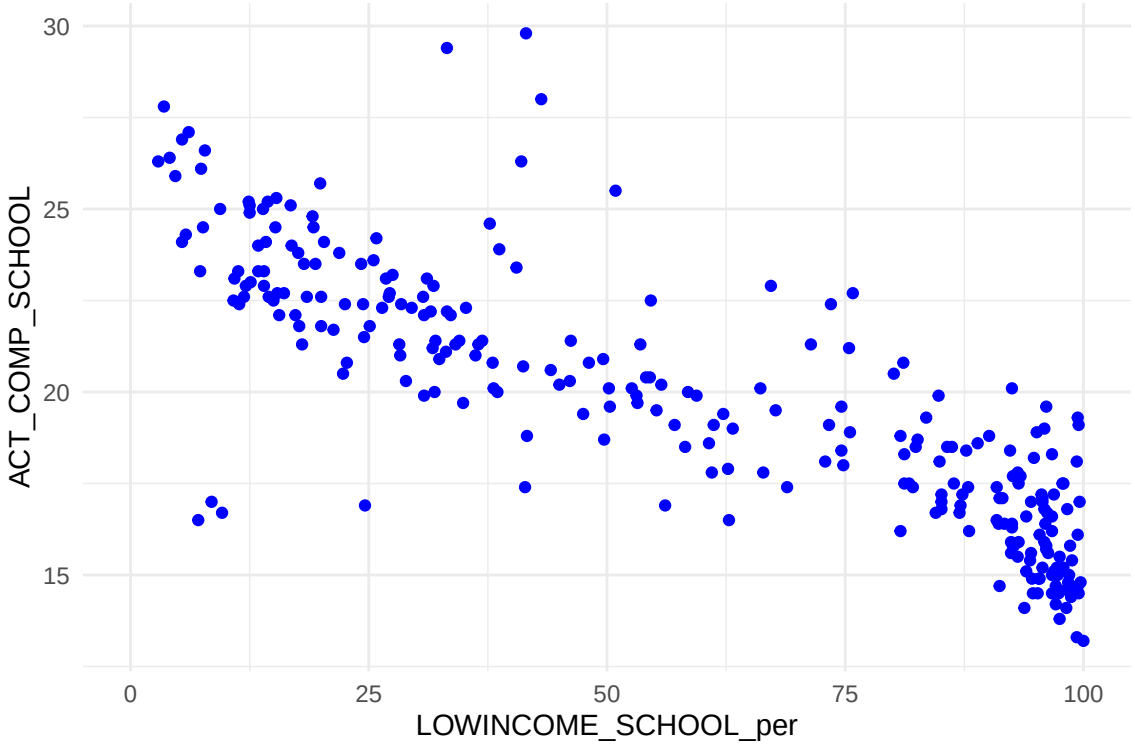
Suppose you choose 3 variables (Composite ACT Score, Enrollment and Attendance Rate) for all schools in the Six County region. What question(s) and display(s) would you explore?

## 7.3 Scatterplot Analysis

### 7.3.1 Outliers and Influential

```
ACTvLI <- sixco %>%
  ggplot(mapping = aes(x = LOWINCOME_SCHOOL_per, y = ACT_COMP_SCHOOL)) +
  geom_point(color="Blue") +
  theme_minimal()
ACTvLI
```

```
## Warning: Removed 1779 rows containing missing values (geom_point).
```

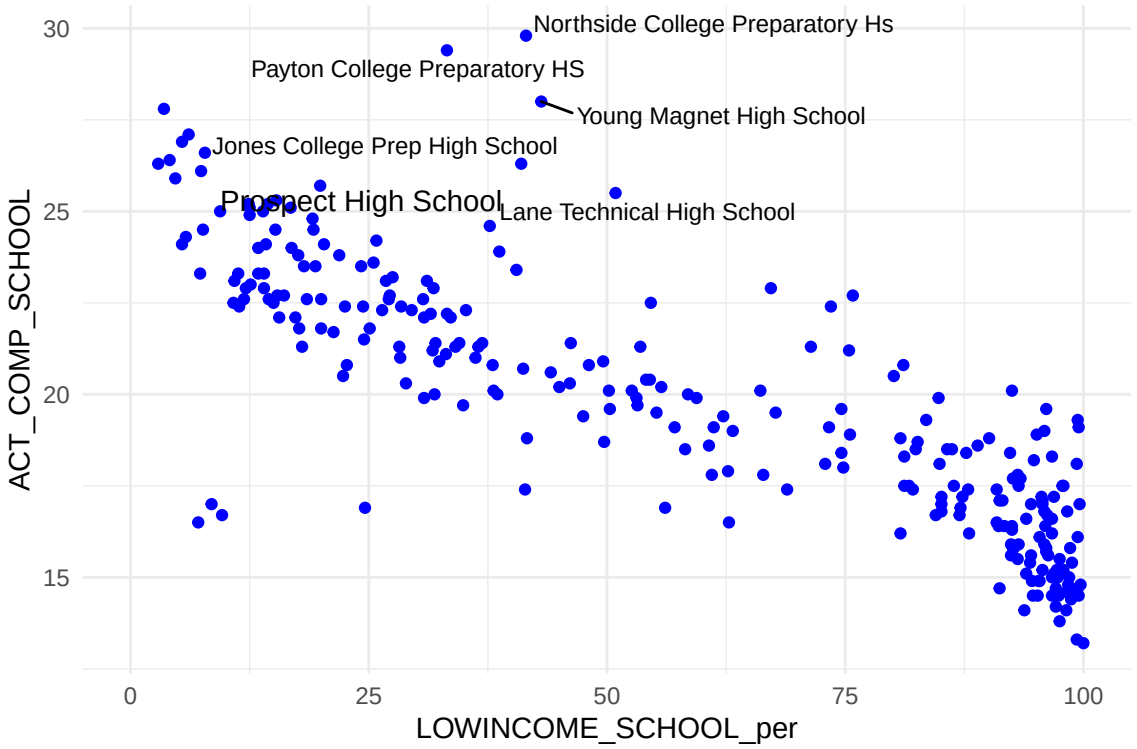



Add some labels using the `ggrepel` package. (Also notice the layering of information.)

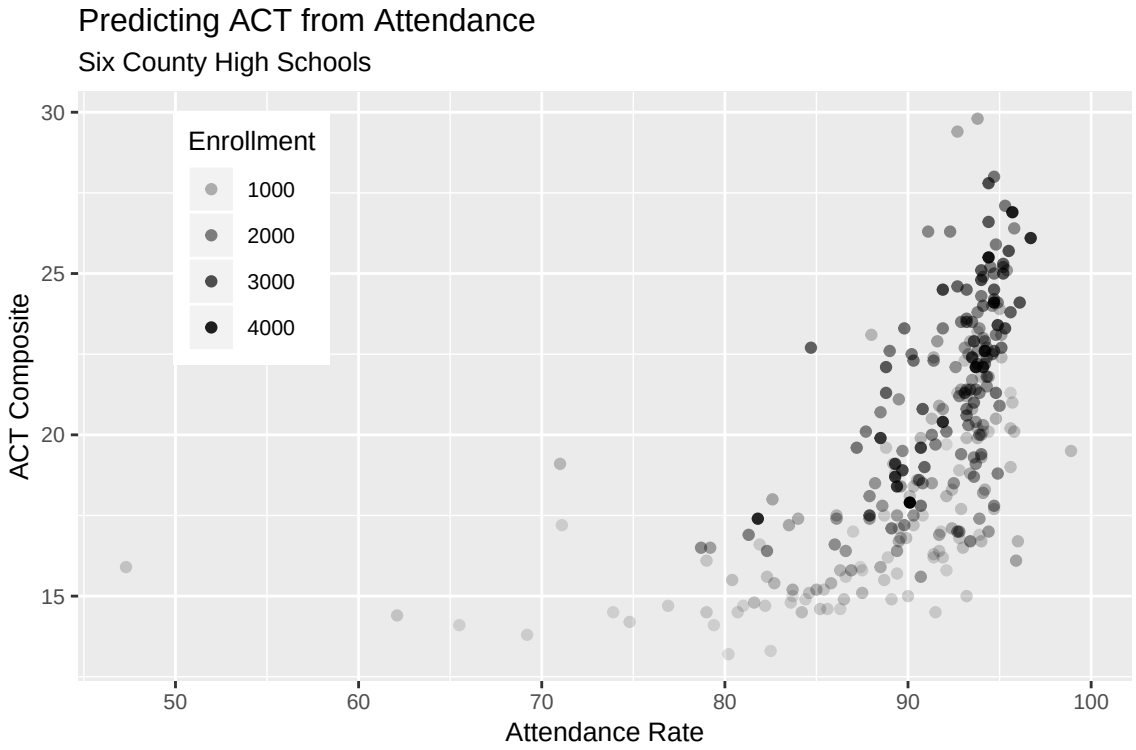
```
library(ggrepel)
```

```
ACTvLI +
  geom_text(aes(label=ifelse(SCHOOL_ID =="050162140170005",
                             as.character(SCHOOL_NAME),'')),hjust=0,vjust=0) +
  geom_text_repel(aes(LOWINCOME_SCHOOL_per, ACT_COMP_SCHOOL,
                      label = ifelse(ACT_COMP_SCHOOL >25 &
                                       LOWINCOME_SCHOOL_per>25,
                                       SCHOOL_NAME, "")), size = 3)
```

```
## Warning: Removed 1779 rows containing missing values (geom_point).
## Warning: Removed 1779 rows containing missing values (geom_text).
## Warning: Removed 1779 rows containing missing values (geom_text_repel).
```




### 7.3.2 Adding a Third Variable




### 7.3.3 Regression in Fathom

Force a variable to be numeric (option) or categorical (shift). Adding a Third Variable: Are charter schools different?

## 7.4 Regression Output

### 7.4.1 All Schools

```
summary(lm(sixco$ACT_COMP_SCHOOL~sixco$ATTENDANCE_RATE_SCHOOL_perALL))
```

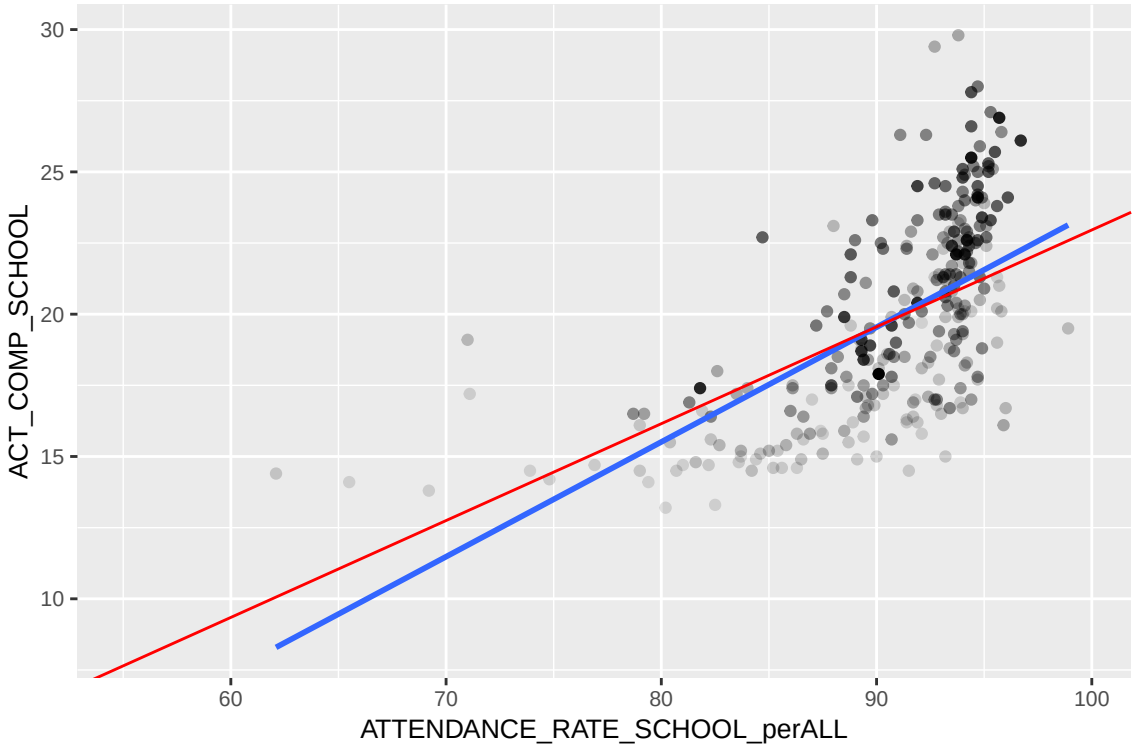
```
##
## Call:
## lm(formula = sixco$ACT_COMP_SCHOOL ~ sixco$ATTENDANCE_RATE_SCHOOL_perALL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7088 -2.2108 -0.2343  1.8209 10.9398
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       -11.26875    2.61192  -4.314 2.25e-05
## sixco$ATTENDANCE_RATE_SCHOOL_perALL   0.34311    0.02887  11.886  < 2e-16
##
## (Intercept)                       ***
## sixco$ATTENDANCE_RATE_SCHOOL_perALL ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.844 on 268 degrees of freedom
##   (1779 observations deleted due to missingness)
## Multiple R-squared:  0.3452, Adjusted R-squared:  0.3427
## F-statistic: 141.3 on 1 and 268 DF,  p-value: < 2.2e-16
```

### 7.4.2 Influential Removed

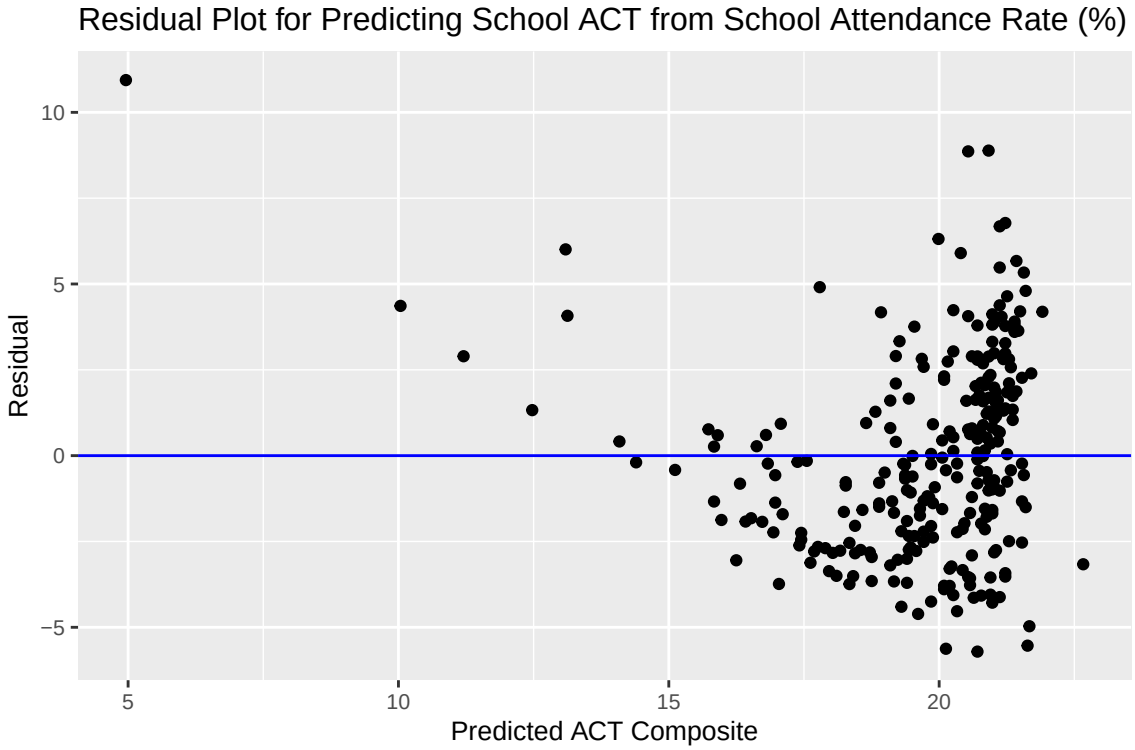
```
sixco_removed <- sixco %>%
  filter(ATTENDANCE_RATE_SCHOOL_perALL>50)

  summary(lm(sixco_removed$ACT_COMP_SCHOOL~sixco_removed$ATTENDANCE_RATE_SCHOOL_perALL))
```

```
##
## Call:
## lm(formula = sixco_removed$ACT_COMP_SCHOOL ~ sixco_removed$ATTENDANCE_RATE_SCHOOL_perALL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8340 -2.1147 -0.1728  1.6627  8.7676
##
## Coefficients:
##                                           Estimate Std. Error t value
## (Intercept)                              -16.74644    2.81182  -5.956
## sixco_removed$ATTENDANCE_RATE_SCHOOL_perALL   0.40322    0.03104  12.993
##                                          Pr(>|t|)
## (Intercept)                              8.13e-09 ***
## sixco_removed$ATTENDANCE_RATE_SCHOOL_perALL  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.75 on 267 degrees of freedom
##   (1778 observations deleted due to missingness)
## Multiple R-squared:  0.3873, Adjusted R-squared:  0.385
## F-statistic: 168.8 on 1 and 267 DF,  p-value: < 2.2e-16
```

### 7.4.3 Both Models




### 7.4.4 Residual Plot



# 8 Random Selection and Simulation

## 8.1 Rolling a die

```
set.seed(2020)
one_die <- sample(1:6, 10, replace = TRUE)
one_die
```

```
##  [1] 4 4 6 1 1 4 2 6 1 5
```

## 8.2 Random Selection

```
four_schools <-  sample_n(rc17, 4)
four_schools[c("SCHOOL_NAME", "SCHOOL_TOTAL_ENROLLMENT")]
```

```
## # A tibble: 4 x 2
##   SCHOOL_NAME                   SCHOOL_TOTAL_ENROLLMENT
##   <chr>                                          <dbl>
## 1 South Side Elementary School                     312
## 2 Coventry Elem School                             565
## 3 Plano High School                                714
## 4 Wanda Kendall Elem School                        236
```

```
mean(four_schools$SCHOOL_TOTAL_ENROLLMENT)
```

```
## [1] 456.75
```

## 8.3 Stratified Sample

```
strat_samp <- sixco %>%
  filter(SCHOOL_TYPE_NAME == "HIGH SCHOOL") %>%
  group_by(COUNTY) %>%
  sample_n(3)
strat_samp[c("SCHOOL_NAME", "COUNTY", "SCHOOL_TOTAL_ENROLLMENT")]
```

```
## # A tibble: 18 x 3
## # Groups:   COUNTY [6]
##    SCHOOL_NAME                       COUNTY  SCHOOL_TOTAL_ENROLLMENT
##    <chr>                             <chr>                     <dbl>
##  1 Marshall Metropolitan High School Cook                        358
##  2 Proviso West High School          Cook                       1850
##  3 Chicago Vocational Career Acad HS Cook                        901
##  4 Westmont High School              Dupage                      449
##  5 Glenbard East High School         Dupage                     2244
##  6 Hinsdale South High School        Dupage                     1507
##  7 Kaneland Senior High School       Kane                       1342
##  8 East High School                  Kane                       3848
##  9 St Charles North High School      Kane                       1985
## 10 Highland Park High School         Lake                       2040
## 11 Libertyville High School          Lake                       1935
## 12 Zion-Benton Twnshp Hi Sch         Lake                       2263
## 13 Woodstock North High School       McHenry                     942
## 14 Huntley High School               McHenry                    2996
## 15 Crystal Lake South High School    McHenry                    1527
## 16 Bolingbrook High School           Will                       3469
## 17 Lincoln-Way Central High School   Will                       2157
## 18 Wilmington High School            Will                        465
```

## 8.4 Confidence Interval Simulation

INCLUDE SD IN TABLE SO CAN CALCULATE INTERVAL BY HAND

```
rand_samp <- function(samples, vari, samp_size) {
  sixco_hs_nona <- sixco_hs[!is.na(sixco_hs[vari]), ] #remove schools with no value
  a <- matrix(ncol = 6, nrow = samples)
  for (k in 1:samples){
    dat_fra <- sample_n(sixco_hs_nona, samp_size)
    t_star <- qt(.975, df = samp_size - 1)
    x_bar <- mean(dat_fra[[vari]])
    stan_dev <- sd(dat_fra[[vari]])
    lower_b <- x_bar  - t_star*stan_dev/(samp_size)**.5
    upper_b <- x_bar + t_star*stan_dev/(samp_size)**.5
  v1 <- k
  v2 <- x_bar
  v3 <- lower_b
  v4 <- upper_b
  v5 <- mean(sixco_hs[[vari]], na.rm = TRUE)
  v6 <- samp_size
  a[k,] <- c(v1, v2, v3, v4, v5, v6)}
  colnames(a) <- c("Sample", "Mean", "L_Bound", "U_Bound", "Parameter", "Sample_Size")
  return(a)
  }
```

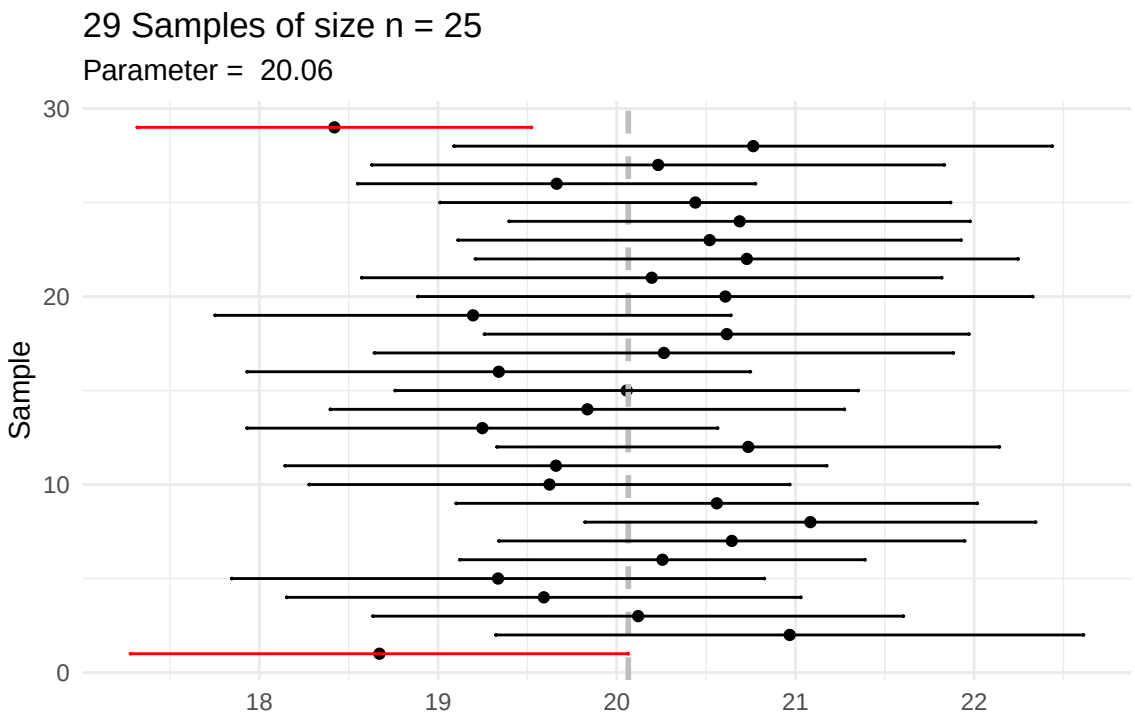
sum(dat_fra[[vari]] !="", na.rm = TRUE)

```
confid_ints <- as_tibble(rand_samp(29, "ACT_COMP_SCHOOL", 25))
confid_ints
```

```
## # A tibble: 29 x 6
##    Sample  Mean L_Bound U_Bound Parameter Sample_Size
##     <dbl> <dbl>   <dbl>   <dbl>     <dbl>       <dbl>
##  1      1  18.7    17.3    20.1      20.1          25
##  2      2  21.0    19.3    22.6      20.1          25
##  3      3  20.1    18.6    21.6      20.1          25
##  4      4  19.6    18.2    21.0      20.1          25
##  5      5  19.3    17.8    20.8      20.1          25
##  6      6  20.3    19.1    21.4      20.1          25
##  7      7  20.6    19.3    21.9      20.1          25
##  8      8  21.1    19.8    22.3      20.1          25
##  9      9  20.6    19.1    22.0      20.1          25
## 10     10  19.6    18.3    21.0      20.1          25
## # ... with 19 more rows
```

```
ggplot(confid_ints, mapping = aes(x=L_Bound, xend = U_Bound, y = Sample)) +
  geom_point(aes(x=Mean, y=Sample)) +
    geom_vline(xintercept = mean(confid_ints$Parameter),
               linetype="dashed",
                color = "grey",
               size=1) +
    geom_dumbbell(size_xend=0,size_x=0,
                color = ifelse(confid_ints$U_Bound < confid_ints$Parameter |
                                 confid_ints$L_Bound > confid_ints$Parameter,
                                 "red", "black")) +
    labs(x = "",
         title = paste(max(confid_ints$Sample),
                       "Samples of size n =",
                       max(confid_ints$Sample_Size)
```

```
                    ),
        subtitle = paste("Parameter = ",round(confid_ints$Parameter,2))
        ) +
  theme_minimal()
```




## 8.5 Binomial

A basketball player claims he makes 70% of his free throws. During a recent game he made only 4 of 10. Does this cast doubt on his 70% claim or could making only 4 of 10 happen to a 70% shooter?

```
free_throws <- rbinom(300, 10, .7)
free_throws
```

```
##   [1]  7  7  7  9  8  8  6  6  8  9  6  5  8  7  6  8  7  7  7  8  7  9  8
##  [24]  6  8  7  8  4  7  6  9  5  5  8  6  7  7  6  8  7  9  9  5  9  8  8
##  [47]  8  8  6  9  8  7  7  5  7  8  8  9  7  6  8  5  7  4  7  6  5  9  7
##  [70]  6  7  5  5  7  8  9  5  6  6  7  9  7  8  6  8  9  7  7  9  7  7  9
##  [93]  8  9  7  8  8  7  7  6 10  6  5  5  7  9  8  8  6  9  7  8  7  6  7
## [116]  9  6  7  5  8  7  8  9  5  7  5  8  8  5 10  8  6  6  7  9  9  7  7
## [139]  8  7  8  5  7  8  4  7  5  7  7  9  5  7  7  8  5  8  8  8  8  5 10
## [162]  8  9  8  6  8 10  6  5  4  7  6  6  9  8  4  7  5  7  8  7  7  7  7
## [185]  7  7  7  6  6  9  8  7  7  8  8  9  9  7  9  6  6  6  6  6  7 10  5
## [208]  8  5  6  6  6  6  9  8  7  9  7  7  4  8  6  8  9  7  8  8  7  8  7
## [231]  7  6  7  8  4  5  8  7  8  7  7  7  6  5  7  6  6  6  8  7  7  7  9
## [254]  8  6  6  8  6  5  6  6  8  7 10  7  9  9  7  9  5  8  6  7  6  6  7
## [277]  5  9  7  8  6  5  3  8  8  8  8  7  6  4  6  8  6  5  9  8  9  8  3
## [300]  7
```

```
sort(free_throws)
```

```
##   [1]  3  3  4  4  4  4  4  4  4  4  5  5  5  5  5  5  5  5  5  5  5  5  5
##  [24]  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  6  6  6  6
##  [47]  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6
##  [70]  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6
##  [93]  6  6  6  6  6  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
```

```
## [116]  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
## [139]  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
## [162]  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  8
## [185]  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
## [208]  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
## [231]  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
## [254]  8  8  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9
## [277]  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9 10 10 10 10 10
## [300] 10
```

```
table(free_throws)
```

```
## free_throws
##  3  4  5  6  7  8  9 10
##  2  8 32 55 86 72 39  6
```

# 9 Appendix

## 9.1 Learning More

### 9.1.1 R and RStudio

R4DS ### Fathom ### Statistics DePaul, Udacity, CodeAcademy, DataCamp Generic Graphs