> # Importing GSS 2006 Data

> GSS2006 <- read\_csv("GSS\_2006.csv")

Parsed with column specification:

cols(

.default = col\_character(),

prestg10 = col\_integer(),

prestg105plus = col\_integer(),

sppres10 = col\_integer(),

sppres105plus = col\_integer(),

papres10 = col\_integer(),

papres105plus = col\_integer(),

mapres10 = col\_integer(),

mapres105plus = col\_integer(),

sei10 = col\_double(),

spsei10 = col\_double(),

pasei10 = col\_double(),

masei10 = col\_double(),

sei10educ = col\_double(),

spsei10educ = col\_double(),

pasei10educ = col\_double(),

masei10educ = col\_double(),

sei10inc = col\_double(),

spsei10inc = col\_double(),

pasei10inc = col\_double(),

masei10inc = col\_double()

# ... with 110 more columns

)

See spec(...) for full column specifications.

Warning: 7 parsing failures.

row # A tibble: 5 x 5 col row col expected actual file expected <int> <chr> <chr> <chr> <chr> actual 1 1722 physhlth an integer DONT KNOW 'GSS\_2006.csv' file 2 2958 prozfor1 an integer DONT KNOW 'GSS\_2006.csv' row 3 3164 adults no trailing characters " or more" 'GSS\_2006.csv' col 4 3170 sphrs2 no trailing characters + hrs 'GSS\_2006.csv' expected 5 3246 physhlth an integer DONT KNOW 'GSS\_2006.csv'

... ................................. ... ................................................................. ........ ....................................... [... truncated]

Warning message:

In rbind(names(probs), probs\_f) :

number of columns of result is not a multiple of vector length (arg 1)

> deg\_child\_06 <- GSS2006 %>%

+ select(degree, childs) %>%

+ filter(!is.na(degree)) %>%

+ filter(!is.na(childs))

Warning message:

package ‘bindrcpp’ was built under R version 3.4.4

> count(deg\_child\_06, degree) %>%

+ mutate(percent = n/sum(n)\*100) %>%

+ arrange(desc(percent)) %>%

+ mutate(cumulative\_prc = cumsum(percent))

# A tibble: 5 x 4

degree n percent cumulative\_prc

<chr> <int> <dbl> <dbl>

1 high school 2269 50.5 50.5

2 bachelor 760 16.9 67.4

3 lt high school 687 15.3 82.7

4 graduate 403 8.96 91.6

5 junior college 377 8.39 100

> #unique responses

> unique(deg\_child\_06$childs)

[1] "3" "1" "2" "0"

[5] "4" "7" "eight or more" "5"

[9] "6"

> deg\_child\_06$childs <- recode(deg\_child\_06$childs,

+ "0" = 0,

+ "1" = 1,

+ "2" = 2,

+ "3" = 3,

+ "4" = 4,

+ "5" = 5,

+ "6" = 6,

+ "7" = 7,

+ "eight or more" = 8

+ )

> count(deg\_child\_06, childs) %>%

+ mutate(percent = n/sum(n)\*100) %>%

+ arrange(desc(percent)) %>%

+ mutate(cumulative\_prc = cumsum(percent))

# A tibble: 9 x 4

childs n percent cumulative\_prc

<dbl> <int> <dbl> <dbl>

1 0 1216 27.0 27.0

2 2 1147 25.5 52.6

3 3 738 16.4 69.0

4 1 710 15.8 84.8

5 4 385 8.56 93.3

6 5 139 3.09 96.4

7 6 83 1.85 98.3

8 7 43 0.956 99.2

9 8 35 0.778 100

> # Table for 'degree' and 'childs'

> table(deg\_child\_06$degree, deg\_child\_06$childs)

0 1 2 3 4 5 6 7 8

bachelor 281 107 189 115 41 16 9 2 0

graduate 156 66 112 43 14 4 6 1 1

high school 563 383 599 400 209 54 37 15 9

junior college 100 64 111 54 27 13 4 2 2

lt high school 116 90 136 126 94 52 27 23 23

> deg\_child\_06 %>%

+ group\_by(degree) %>%

+ summarize(count= n(), mean = mean(childs), std.dev = sd(childs)) %>%

+ arrange(desc(mean)) %>%

+ print()

# A tibble: 5 x 4

degree count mean std.dev

<chr> <int> <dbl> <dbl>

1 lt high school 687 2.74 2.08

2 high school 2269 1.89 1.57

3 junior college 377 1.79 1.56

4 bachelor 760 1.50 1.48

5 graduate 403 1.35 1.42

> # ANOVA test

> model <- aov(childs ~ degree, data = deg\_child\_06)

> summary(model) # getting the summary of our ANOVA model

Df Sum Sq Mean Sq F value Pr(>F)

degree 4 730 182.59 68.4 <2e-16 \*\*\*

Residuals 4491 11989 2.67

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1