STAT 225 - Prelim Project Look Group X

Names: Harrison Marick, Chase Henley, Joe Feldman (Group D)

What is the True Value of a College Education?

Perhaps the most prevalent rationale for the pursuance of a college degree is the promise of economic mobility upon graduation. However, our project aims to investigate whether this opportunity is equally available for all students. Specifically, our group will analyze whether the racial and socio-demographic composition of a college affects the proportion of students who go on to earn higher wages after enrolling in the institution.

Read in the data

```
data<-read.csv("Scorecard.csv")
data2<-data[,c(-6, -18, -19, -(27:34), -(42:89))]
data2<-data2[,c(-1,-2,-3,-5)]#simplifying and only selecting certain columns# Must start with a data-re
```

Summary command on data set

```
summary(data2)
```

```
##                                     INSTNM                 PCT_WHITE
##  McCann School of Business & Technology:   9   NULL              :2166
##  Stevens-Henager College                :   7   PrivacySuppressed:  95
##  Bryan University                       :   5   87.5999984741211 :  24
##  Columbia College                       :   5   69.4700012207031 :  17
##  Unitek College                         :   4   75.0100021362304 :  17
##  Wentworth Institute of Technology      :   4   62.7999992370605 :  15
##  (Other)                                :7559   (Other)          :5259
##               PCT_BLACK                PCT_ASIAN
##  NULL              :2166   NULL              :2166
##  PrivacySuppressed:  95   PrivacySuppressed:  95
##  7.34000015258789 :  24   0.47999998927116 :  33
##  7.09000015258789 :  18   2.23000001907348 :  33
##  15.369999885559  :  17   0.20000000298023 :  31
##  17.8799991607666 :  15   0.87000000476837 :  29
##  (Other)          :5258   (Other)          :5206
##             PCT_HISPANIC                PCT_BA
##  NULL              :2166   NULL              :2166
##  PrivacySuppressed:  95   PrivacySuppressed: 133
##  3.47000002861023 :  27   15.4099998474121 :  26
##  1.41999995708465 :  20   12.1800003051757 :  20
##  4.78999996185302 :  20   12.3500003814697 :  20
##  16.2299995422363 :  17   15.3900003433227 :  20
##  (Other)          :5248   (Other)          :5208
##            PCT_GRAD_PROF             PCT_BORN_US
##  NULL              :2166   NULL              :2166
##  PrivacySuppressed: 133   PrivacySuppressed: 133
##  9.46000003814697 :  28   93.4300003051757 :  25
##  5.61999988555908 :  25   86.6999969482422 :  17
##  7.5              :  25   93.1100006103515 :  17
##  7.78999996185302 :  23   93.1600036621093 :  16
```

```
##  (Other)          :5193   (Other)          :5219
##            MEDIAN_HH_INC             POVERTY_RATE
##  NULL             :2166   NULL             :2166
##  PrivacySuppressed:  95   PrivacySuppressed:  95
##  63381.69         :  23   6.78999996185302 :  35
##  62086.89         :  17   7.1100001335144  :  22
##  48773.91         :  14   7.98000001907348 :  21
##  55670.96         :  13   9.92000007629394 :  21
##  (Other)          :5265   (Other)          :5233
##              UNEMP_RATE           LN_MEDIAN_HH_INC
##  NULL             :2166   NULL            :2166
##  PrivacySuppressed:  95   10.9799995422363: 122
##  3.15000009536743 :  50   10.8900003433227: 120
##  3.16000008583068 :  48   10.9300003051757: 120
##  3.80999994277954 :  47   10.9700002670288: 116
##  3.03999996185302 :  41   10.8800001144409: 113
##  (Other)          :5146   (Other)         :4836
##           MN_EARN_WNE_P10          MD_EARN_WNE_P10
##  NULL             :1139   NULL             :1139
##  PrivacySuppressed: 772   PrivacySuppressed: 772
##  43100            : 141   38800            : 148
##  24400            :  87   21800            :  87
##  54100            :  84   47300            :  87
##  32400            :  48   29600            :  62
##  (Other)          :5322   (Other)          :5298
##          PCT10_EARN_WNE_P10         PCT25_EARN_WNE_P10
##  NULL             :1625     PrivacySuppressed:1273
##  PrivacySuppressed:1261     NULL             :1139
##  8900             : 165     23000            : 148
##  4400             : 130     11700            : 117
##  6200             :  99     28100            :  95
##  11500            :  85     13500            :  61
##  (Other)          :4228     (Other)          :4760
##          PCT75_EARN_WNE_P10         PCT90_EARN_WNE_P10
##  PrivacySuppressed:1273     NULL             :1625
##  NULL             :1139     PrivacySuppressed:1261
##  57200            : 139     77000            : 128
##  33800            :  88     42700            :  78
##  73100            :  81     106500           :  77
##  44200            :  49     53800            :  34
##  (Other)          :4824     (Other)          :4390
##           SD_EARN_WNE_P10             GT_25K_P10
##  NULL             :1139   NULL             :1139
##  PrivacySuppressed: 772   PrivacySuppressed: 772
##  29800            : 180   0.722            : 142
##  16900            :  91   0.785            :  89
##  38700            :  84   0.423            :  84
##  19500            :  56   0.594            :  46
##  (Other)          :5271   (Other)          :5321
##         MN_EARN_WNE_INC1_P10        MN_EARN_WNE_INC2_P10
##  PrivacySuppressed:1810      PrivacySuppressed:3042
##  NULL             :1139      NULL             :1139
##  38400            : 141      47300            : 147
##  41600            : 110      53600            :  82
```

```
##  22400            :  86      24200            :  81
##  29700            :  58      34200            :  50
##  (Other)          :4249      (Other)          :3052
##         MN_EARN_WNE_INC3_P10    MN_EARN_WNE_INDEP0_INC1_P10
##  PrivacySuppressed:3031      PrivacySuppressed:3391
##  NULL             :1139      NULL             :1625
##  47000            : 148      38500            : 130
##  72700            :  82      40000            :  78
##  27400            :  81      26500            :  37
##  46800            :  59      37800            :  36
##  (Other)          :3053      (Other)          :2296
##        MN_EARN_WNE_INDEP0_P10       MN_EARN_WNE_INDEP1_P10
##  PrivacySuppressed:2187       PrivacySuppressed:2187
##  NULL             :1139       NULL             :1139
##  40500            : 146       45100            : 141
##  46200            :  90       24200            :  90
##  24500            :  82       54800            :  85
##  32000            :  48       32100            :  51
##  (Other)          :3901       (Other)          :3900
```

Data Codebook

List your variables and whether they are quantitative/qualitative (numeric vs categorical), along with other notes about the variables. Hint, to make a nice list, you need to put two spaces at the end of each line to force RMarkdown to start a new line. Check this out below:

Variable 1 - INSTM: Institution Name; qualitative

Variable 2 - PCT_WHITE: Percent of the population from students' zip codes that is White, via Census data; quantitative

Variable 3 -
Variable 4 - Variable 5 - Variable 6 - Variable 7 - Variable 8 - Variable 9 - Variable 10 - Variable 11- Variable 12- Variable 13- Variable 14- Variable 15- Variable 16- Variable 17- Variable 18- Variable 19- SD_EARN_WNE_P10: Standard deviation of earnings of students working and not enrolled 10 years after entry; quantitative Variable 20- GT_25K_P10: Share of students earning over $25,000/year (threshold earnings) 10 years after entry; quantitative Variable 21- MN_EARN_WNE_INC1_P10: Mean earnings of students working and not enrolled 10 years after entry in the lowest income tercile $0-30,000; quantitative Variable 22- MN_EARN_WNE_INC2_P10: Mean earnings of students working and not enrolled 10 years after entry in the middle income tercile 30,001-75,000 dollars; quantitative Variable 23- MN_EARN_WNE_INC3_P10: Mean earnings of students working and not enrolled 10 years after entry in the highest income tercile $75,001+ ; quantitative Variable 24- MN_EARN_WNE_INDEP0_INC1_P10: Mean earnings of dependent students working and not enrolled 10 years after entry in the lowest income tercile 0-30,000 dollars; quantitative Variable 25- MN_EARN_WNE_INDEP0_P10: Mean earnings of dependent students working and not enrolled 10 years after entry; quantitative Variable 26- MN_EARN_WNE_INDEP1_P10: Mean earnings of independent students working and not enrolled 10 years after entry; quantitative ### Analysis Plan

A rough outline for your proposed analysis, including univariate summaries, bivariate relationships, and the like

Prelim Univariate Analysis

Obtain basic univariate descriptive statistics and graphs for variables relevant to your analysis.

```
#for quantitative variables, you could use
#histogram, bwplot, densityplot
#favstats
```

```
# for qualitative variables, you could use
#tally
#bargraph
```

COMMENT on what you see!

Prelim Bivariate Analysis

Make at least one scatterplot or side-by-side boxplot to examine a bivariate relationship where the response is quantitative (needed for chapter 6, or 8-9 analysis).

```
#bwplot or xyplot
```

COMMENT on what you see!

Any other Analysis Thoughts

Questions for me

Joe: codebook for variables 1-9. One univariate analysis (proportion earning over 25k). One bivariate (% white as explanation of something)

Harrison: codebook for variables 10-18. One univariate analysis (distribution of whiteness). One bivariate (student earnings 10 years post by whiteness)

Chase: codebook for variables 19-26. One univariate analysis (median earnings 10 years post).