

mbruner3_1.r

Mark Bruner

9/11/2020

Reference

Data is taken from <http://faculty.marshall.usc.edu/gareth-james/ISL/College.csv>

Start

Loaded in the college.csv and assigned it to the variable “college”. Renamed the first column in the data set.

```
library(readr)
library(dplyr)
college <- read_csv("college.csv", col_types = "cciiiiiiiiiiiiiiii")
college <- rename(college, "College Name" = "X1")
college
```

```
## # A tibble: 777 x 19
##   'College Name' Private  Apps Accept Enroll Top10perc Top25perc F.Undergrad
##   <chr>           <chr>   <int> <int>   <int>    <int>    <int>    <int>
## 1 Abilene Chris~ Yes     1660  1232   721      23      52     2885
## 2 Adelphi Unive~ Yes     2186  1924   512      16      29     2683
## 3 Adrian College Yes     1428  1097   336      22      50     1036
## 4 Agnes Scott C~ Yes      417   349   137      60      89      510
## 5 Alaska Pacifi~ Yes      193   146    55      16      44      249
## 6 Albertson Col~ Yes      587   479   158      38      62      678
## 7 Albertus Magn~ Yes      353   340   103      17      45      416
## 8 Albion College Yes     1899  1720   489      37      68     1594
## 9 Albright Coll~ Yes     1038   839   227      30      63      973
## 10 Alderson-Broa~ Yes      582   498   172      21      44      799
## # ... with 767 more rows, and 11 more variables: P.Undergrad <int>,
## #   Outstate <int>, Room.Board <int>, Books <int>, Personal <int>, PhD <int>,
## #   Terminal <int>, S.F.Ratio <int>, perc.alumni <int>, Expend <int>,
## #   Grad.Rate <int>
```

Loaded “summarytools” package.

```
library(summarytools)
```

Quantitative Descriptive Statistics

Made the “Apps” column in college data set a variable and ran summary of desc. statistics.

```
apps <- college$Apps
descr(x = apps)
```

```
## Descriptive Statistics
## apps
## N: 777
##
##           apps
## -----
##           Mean    3001.64
##           Std.Dev  3870.20
##           Min      81.00
##           Q1       776.00
##           Median   1558.00
##           Q3       3624.00
##           Max      48094.00
##           MAD       1463.33
##           IQR       2848.00
##           CV        1.29
##           Skewness   3.71
##           SE.Skewness 0.09
##           Kurtosis   26.52
##           N.Valid    777.00
##           Pct.Valid  100.00
```

##Qualitative Descriptive Statistics & Transformation of Data Used the “Private” column in the College data set and converted it to a DF. After that I added a column to the DF to represent the percentage.

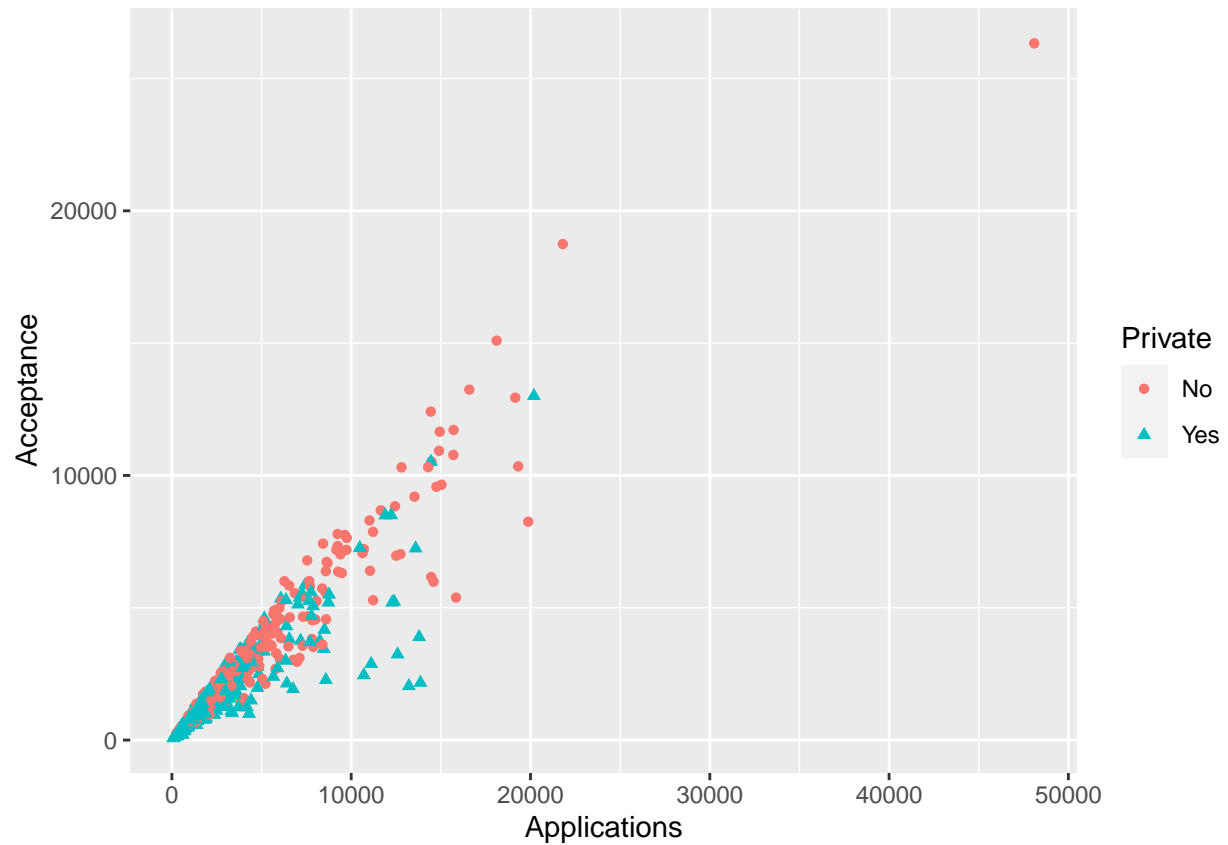
```
library(tidyverse)
private <- table(college$Private)
private_df <- as.data.frame(private)
private_df <- rename(private_df, "Private" = "Var1")
private_percent <- private_df$Freq/sum(private_df$Freq)
private_df_percent <- add_column(private_df, private_percent)
private_df_percent <- rename(private_df_percent, "Percent" = "private_percent")
private_df_percent
```

```
##   Private Freq  Percent
## 1      No   212 0.2728443
## 2     Yes   565 0.7271557
```

Scatterplot: Multi-variable

Created a scatterplot of the two variable Applications and Acceptance in the College Data Set.

```
library(ggplot2)
Applications = college$Apps
Acceptance = college$Accept
Private = college$Private
ggplot(data = college, aes(x = Applications, y = Acceptance, col = Private, shape = Private)) + geom_point()
```



Single Variable Visualization

```
ggplot(data = college, aes(x = Private)) + geom_histogram(stat = "Count", fill = "light blue", col = "skyblue")
```

```
ggplot(data = college, aes(x = Private)) + geom_histogram(stat = "Count", fill = "light blue", col = "skyblue")
```

