Set wokring directory.

setwd("C:\\Users\\Juan Nunez\\Desktop\\DATA\_Capstone")

Open dataset.

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.0 v purrr 0.3.0   
## v tibble 2.0.1 v dplyr 0.8.0.1  
## v tidyr 0.8.2 v stringr 1.4.0   
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

TA\_df <- read\_csv("DATASETS\\Tuition\_Assistance.csv")

## Parsed with column specification:  
## cols(  
## Department = col\_character(),  
## Major = col\_character(),  
## Degree = col\_character(),  
## School = col\_character(),  
## `Course Title` = col\_character(),  
## `Course Description` = col\_character(),  
## Cost = col\_number()  
## )

Look at the top of the dataset.

dim(TA\_df)

## [1] 3093 7

head(TA\_df)

## # A tibble: 6 x 7  
## Department Major Degree School `Course Title` `Course Descript~ Cost  
## <chr> <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 Police Busin~ AA Montgo~ INTRODUCTION ~ An introductory ~ 392  
## 2 Police Busin~ AA Montgo~ MA 160 A general calcul~ 392  
## 3 Police Busin~ AA Montgo~ INTRO TO AMER~ A survey of Amer~ 392  
## 4 Health & H~ Busin~ Master~ BOWIE ~ PUBLIC POLICY~ Focus is on the ~ 1062  
## 5 Health & H~ Other~ Master~ Mount ~ MHA 500 CONTE~ "The historical,~ 1569  
## 6 Finance Accou~ Master~ Univer~ ACCT FINANCIA~ A synthesis of m~ 870

Below we see which departments are recipients of tuition assistance.

library("Hmisc")

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':  
##   
## src, summarize

## The following objects are masked from 'package:base':  
##   
## format.pval, units

unique(TA\_df$Department)

## [1] "Police"   
## [2] "Health & Human Services"   
## [3] "Finance"   
## [4] "County Attorney"   
## [5] "Community Use Public Facilities"   
## [6] "Public Information"   
## [7] "State's Attorney"   
## [8] "General Services"   
## [9] "Fire/Rescue Services"   
## [10] "Correction & Rehabilitation"   
## [11] "Transportation"   
## [12] "Housing & Community Affairs"   
## [13] "County Council"   
## [14] "Technology Services"   
## [15] "Environmental Protection"   
## [16] "Human Resources"   
## [17] "Recreation"   
## [18] "County Executive"   
## [19] "Libraries"   
## [20] "Consumer Protection"   
## [21] "Liquor Control"   
## [22] "Permitting Services"   
## [23] "Investment Trustees"   
## [24] "Sheriff"   
## [25] "Community Engagement Cluster"   
## [26] "Procurement"   
## [27] "Management & Budget"   
## [28] "Board of Elections"   
## [29] "Emergency Mgmt & Homeland Security"  
## [30] "Legislative Oversight"   
## [31] "Human Rights"

table(TA\_df$Department)

##   
## Board of Elections Community Engagement Cluster   
## 6 6   
## Community Use Public Facilities Consumer Protection   
## 25 5   
## Correction & Rehabilitation County Attorney   
## 179 27   
## County Council County Executive   
## 3 1   
## Emergency Mgmt & Homeland Security Environmental Protection   
## 4 30   
## Finance Fire/Rescue Services   
## 63 633   
## General Services Health & Human Services   
## 79 578   
## Housing & Community Affairs Human Resources   
## 50 42   
## Human Rights Investment Trustees   
## 1 33   
## Legislative Oversight Libraries   
## 5 28   
## Liquor Control Management & Budget   
## 61 3   
## Permitting Services Police   
## 14 859   
## Procurement Public Information   
## 15 42   
## Recreation Sheriff   
## 24 12   
## State's Attorney Technology Services   
## 41 38   
## Transportation   
## 186

I create a dummy variable for police.

TA\_df$Police <- ifelse(TA\_df$Department == "Police", 1, 0)  
head(TA\_df)

## # A tibble: 6 x 8  
## Department Major Degree School `Course Title` `Course Descrip~ Cost  
## <chr> <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 Police Busi~ AA Montg~ INTRODUCTION ~ An introductory~ 392  
## 2 Police Busi~ AA Montg~ MA 160 A general calcu~ 392  
## 3 Police Busi~ AA Montg~ INTRO TO AMER~ A survey of Ame~ 392  
## 4 Health & ~ Busi~ Maste~ BOWIE~ PUBLIC POLICY~ Focus is on the~ 1062  
## 5 Health & ~ Othe~ Maste~ Mount~ MHA 500 CONTE~ "The historical~ 1569  
## 6 Finance Acco~ Maste~ Unive~ ACCT FINANCIA~ A synthesis of ~ 870  
## # ... with 1 more variable: Police <dbl>

Let’s see which degrees are pursued.

table(TA\_df$Degree)

##   
## AA Bachelors (BA/BS) Certificate   
## 533 994 333   
## Juris Doctor Masters (MA/MS/MPH/etc.) Non-Degree   
## 26 606 391   
## Other Ph.D. (DCS) Ph.D. (DDE)   
## 160 30 20

I fix the degree variable to make it more intelligible.

TA\_df$Degree\_s[TA\_df$Degree == "Ph.D. (DDE)"] <- "Other"

## Warning: Unknown or uninitialised column: 'Degree\_s'.

TA\_df$Degree\_s[TA\_df$Degree == "Other"] <- "Other"  
TA\_df$Degree\_s[TA\_df$Degree == "Juris Doctor"] <- "Other"  
TA\_df$Degree\_s[TA\_df$Degree == "Ph.D. (DCS)"] <- "Other"  
TA\_df$Degree\_s[TA\_df$Degree == "AA"] <- "AA"  
TA\_df$Degree\_s[TA\_df$Degree == "Masters (MA/MS/MPH/etc.)"] <- "MA"  
TA\_df$Degree\_s[TA\_df$Degree == "Bachelors (BA/BS)"] <- "BA"  
TA\_df$Degree\_s[TA\_df$Degree == "Certificate"] <- "Certificate"  
TA\_df$Degree\_s[TA\_df$Degree == "Non-Degree"] <- "Non-Degree"  
  
table(TA\_df$Degree\_s)

##   
## AA BA Certificate MA Non-Degree Other   
## 533 994 333 606 391 236

head(TA\_df)

## # A tibble: 6 x 9  
## Department Major Degree School `Course Title` `Course Descrip~ Cost  
## <chr> <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 Police Busi~ AA Montg~ INTRODUCTION ~ An introductory~ 392  
## 2 Police Busi~ AA Montg~ MA 160 A general calcu~ 392  
## 3 Police Busi~ AA Montg~ INTRO TO AMER~ A survey of Ame~ 392  
## 4 Health & ~ Busi~ Maste~ BOWIE~ PUBLIC POLICY~ Focus is on the~ 1062  
## 5 Health & ~ Othe~ Maste~ Mount~ MHA 500 CONTE~ "The historical~ 1569  
## 6 Finance Acco~ Maste~ Unive~ ACCT FINANCIA~ A synthesis of ~ 870  
## # ... with 2 more variables: Police <dbl>, Degree\_s <chr>

533 + 994 + 333 + 606 + 391 + 236

## [1] 3093

533 + 994 + 333 + 606 + 391 + 160 + 30 + 26 + 20

## [1] 3093

#TA\_df <- mutate(TA\_df, Police\_cat = as.factor(TA\_df$Police))  
#head(TA\_df)  
#table(TA\_df$Police\_cat)  
  
#png(filename="C:\\Users\\Juan Nunez\\Desktop\\DATA\_Capstone\\midterm\_graph.png")  
  
#ggplot(TA\_df) +  
 # aes(x = Degree\_s, fill = Police\_cat) +  
 #geom\_bar() +  
 #guides(fill=guide\_legend(title="Recipient", labels = "Other", "Police"))  
  
  
#ggplot(data = TA\_df , aes(x=Degree\_s, fill=Police\_cat)) +   
 # geom\_bar() +  
 #labs(title="Tuition Assistance Recipients Counts\n", x="Degree", y="Count") +   
 #guides(fill=guide\_legend(title="Recipient")) +  
 #scale\_fill\_discrete(labels = c("Other", "Police"))  
  
#dev.off()  
  
##factor(XPRESSCAT, labels = c("Completely free", "Very free", "Moderately free", "Slightly free", "Not free at all")))

I fix the department variable.

table(TA\_df$Department)

##   
## Board of Elections Community Engagement Cluster   
## 6 6   
## Community Use Public Facilities Consumer Protection   
## 25 5   
## Correction & Rehabilitation County Attorney   
## 179 27   
## County Council County Executive   
## 3 1   
## Emergency Mgmt & Homeland Security Environmental Protection   
## 4 30   
## Finance Fire/Rescue Services   
## 63 633   
## General Services Health & Human Services   
## 79 578   
## Housing & Community Affairs Human Resources   
## 50 42   
## Human Rights Investment Trustees   
## 1 33   
## Legislative Oversight Libraries   
## 5 28   
## Liquor Control Management & Budget   
## 61 3   
## Permitting Services Police   
## 14 859   
## Procurement Public Information   
## 15 42   
## Recreation Sheriff   
## 24 12   
## State's Attorney Technology Services   
## 41 38   
## Transportation   
## 186

TA\_df$dep\_n[TA\_df$Department == "Police"] <- "Police"

## Warning: Unknown or uninitialised column: 'dep\_n'.

TA\_df$dep\_n[TA\_df$Department == "Fire/Rescue Services"] <- "Fire & Rescue"  
TA\_df$dep\_n[TA\_df$Department == "Health & Human Services"] <- "HHS"  
TA\_df$dep\_n[TA\_df$Department != "Police" & TA\_df$Department != "Fire/Rescue Services" & TA\_df$Department != "Health & Human Services"] <- "Other"  
head(TA\_df)

## # A tibble: 6 x 10  
## Department Major Degree School `Course Title` `Course Descrip~ Cost  
## <chr> <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 Police Busi~ AA Montg~ INTRODUCTION ~ An introductory~ 392  
## 2 Police Busi~ AA Montg~ MA 160 A general calcu~ 392  
## 3 Police Busi~ AA Montg~ INTRO TO AMER~ A survey of Ame~ 392  
## 4 Health & ~ Busi~ Maste~ BOWIE~ PUBLIC POLICY~ Focus is on the~ 1062  
## 5 Health & ~ Othe~ Maste~ Mount~ MHA 500 CONTE~ "The historical~ 1569  
## 6 Finance Acco~ Maste~ Unive~ ACCT FINANCIA~ A synthesis of ~ 870  
## # ... with 3 more variables: Police <dbl>, Degree\_s <chr>, dep\_n <chr>

I create a bar chart with the department and the degree they pursue.

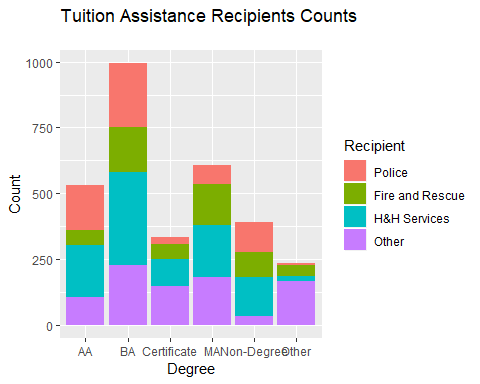
TA\_df <- mutate(TA\_df, deps = as.factor(TA\_df$dep\_n))  
head(TA\_df)

## # A tibble: 6 x 11  
## Department Major Degree School `Course Title` `Course Descrip~ Cost  
## <chr> <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 Police Busi~ AA Montg~ INTRODUCTION ~ An introductory~ 392  
## 2 Police Busi~ AA Montg~ MA 160 A general calcu~ 392  
## 3 Police Busi~ AA Montg~ INTRO TO AMER~ A survey of Ame~ 392  
## 4 Health & ~ Busi~ Maste~ BOWIE~ PUBLIC POLICY~ Focus is on the~ 1062  
## 5 Health & ~ Othe~ Maste~ Mount~ MHA 500 CONTE~ "The historical~ 1569  
## 6 Finance Acco~ Maste~ Unive~ ACCT FINANCIA~ A synthesis of ~ 870  
## # ... with 4 more variables: Police <dbl>, Degree\_s <chr>, dep\_n <chr>,  
## # deps <fct>

table(TA\_df$deps)

##   
## Fire & Rescue HHS Other Police   
## 633 578 1023 859

#png(filename="C:\\Users\\Juan Nunez\\Desktop\\DATA\_Capstone\\midterm\_graph.png")  
  
  
ggplot(data = TA\_df , aes(x=Degree\_s, fill=deps)) +   
 geom\_bar() +  
 labs(title="Tuition Assistance Recipients Counts\n", x="Degree", y="Count") +   
 guides(fill=guide\_legend(title="Recipient")) +  
 scale\_fill\_discrete(labels = c("Police", "Fire and Rescue", "H&H Services", "Other"))



#dev.off()

write.csv(TA\_df, file = "TA\_df.csv",row.names=FALSE, na="")

table(TA\_df$dep\_n)

##   
## Fire & Rescue HHS Other Police   
## 633 578 1023 859

TA\_df$dep\_n <- factor(TA\_df$dep\_n, levels = c("Fire & Rescue", "HHS", "Police", "Other"))  
table(TA\_df$dep\_n)

##   
## Fire & Rescue HHS Police Other   
## 633 578 859 1023

table(TA\_df$Department)

##   
## Board of Elections Community Engagement Cluster   
## 6 6   
## Community Use Public Facilities Consumer Protection   
## 25 5   
## Correction & Rehabilitation County Attorney   
## 179 27   
## County Council County Executive   
## 3 1   
## Emergency Mgmt & Homeland Security Environmental Protection   
## 4 30   
## Finance Fire/Rescue Services   
## 63 633   
## General Services Health & Human Services   
## 79 578   
## Housing & Community Affairs Human Resources   
## 50 42   
## Human Rights Investment Trustees   
## 1 33   
## Legislative Oversight Libraries   
## 5 28   
## Liquor Control Management & Budget   
## 61 3   
## Permitting Services Police   
## 14 859   
## Procurement Public Information   
## 15 42   
## Recreation Sheriff   
## 24 12   
## State's Attorney Technology Services   
## 41 38   
## Transportation   
## 186

library(psych)

## Warning: package 'psych' was built under R version 3.5.3

##   
## Attaching package: 'psych'

## The following object is masked from 'package:Hmisc':  
##   
## describe

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

describe(TA\_df$Cost)

## vars n mean sd median trimmed mad min max range skew  
## X1 1 3093 755.53 522.7 655 702.4 452.19 0 2130 2130 0.83  
## kurtosis se  
## X1 -0.12 9.4

class(TA\_df$Cost)

## [1] "numeric"

table(TA\_df$Degree\_s)

##   
## AA BA Certificate MA Non-Degree Other   
## 533 994 333 606 391 236

TA\_df$Degree\_s <- factor(TA\_df$Degree\_s, levels = c("AA", "BA", "MA", "Certificate","Non-Degree", "Other"))  
table(TA\_df$Degree\_s)

##   
## AA BA MA Certificate Non-Degree Other   
## 533 994 606 333 391 236

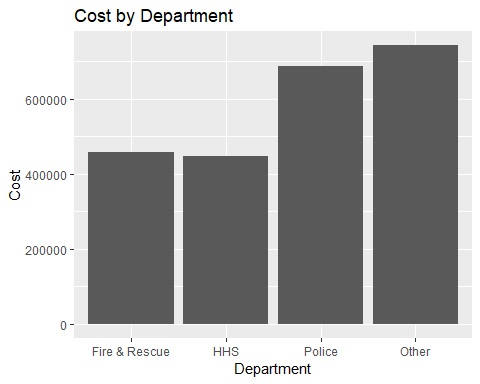
table(TA\_df$Degree)

##   
## AA Bachelors (BA/BS) Certificate   
## 533 994 333   
## Juris Doctor Masters (MA/MS/MPH/etc.) Non-Degree   
## 26 606 391   
## Other Ph.D. (DCS) Ph.D. (DDE)   
## 160 30 20

table(TA\_df$dep\_n)

##   
## Fire & Rescue HHS Police Other   
## 633 578 859 1023

#filter(!is.na(AGE\_GROUP\_fixed)) %>%  
options(scipen=10000)  
TA\_df %>%   
ggplot(aes(y = Cost, x = as.factor(dep\_n))) +  
geom\_bar(stat="identity") +  
labs(title = "Cost by Department", x= "Department", y = "Cost")



options(scipen=10000)  
TA\_df %>%   
ggplot(aes(y = Cost, x = as.factor(Degree\_s))) +  
geom\_bar(stat="identity") +  
labs(title = "Cost by Degree", x= "Degree", y = "Cost")

