DATA VIZ - FINAL PROJECT

NATHANIEL ASIEDU SAKYI*
UNIVERSITY OF TEXAS AT EL PASO (UTEP)

2023-05-12

Contents

---
*nasakyi@miners.utep.edu

1 Data Preparation

1.1 Bringing the Master and Extra Data Sets into R

```
master <- read.csv("master.csv")
extra <- read.csv("extra_questions_withID.csv")
```

1.2 Inspecting the Variables in Both Data

```
str(master)
```

```
## 'data.frame':    11763 obs. of  59 variables:
##  $ X.1          : int  3 4 5 8 9 11 15 16 18 22 ...
##  $ X            : int  3 4 5 8 9 11 15 16 18 22 ...
##  $ RespondentId : int  83001882 83001883 83001889 83001892 83001893 83001897 8300190
##  $ Enroll       : chr  "Part-time" "Full-time" "Full-time" "Part-time" ...
##  $ Employ       : chr  "Full-time" "Full-time" "Full-time" "Part-time" ...
##  $ Working      : chr  "On campus" "On campus" "On campus" NA ...
##  $ Hrs          : chr  "19 hours or less" "19 hours or less" "More than 19 hours" NA
##  $ Age          : int  3 2 3 2 2 3 2 2 4 2 ...
##  $ Ethnicity    : chr  "4" "4" "1" "1" ...
##  $ Income       : chr  "Less than $10,000" "$40,000 to $49,999" "Less than $10,000"
##  $ Classification: chr  "Senior" "Senior" "Doctoral" "Junior" ...
##  $ College      : chr  "5" "1" "3" "7" ...
##  $ Commute      : chr  "Car (someone drives to campus and picks you up)" "Car (someo
##  $ Alone        : chr  "No" "Yes" "Yes" "No" ...
##  $ Dependents   : chr  "None" NA NA "None" ...
##  $ HoH          : chr  "No" "Yes" "Yes" "No" ...
##  $ Live         : chr  "Off campus with parents" "Off campus with partner" "Off camp
##  $ FedAid       : chr  "Loans" "Work-study" "Grants" "Loans" ...
##  $ PA           : chr  "Yes" "No" "Yes" "Yes" ...
##  $ FreqNightElse : chr  NA "Often" NA NA ...
##  $ UTEPHomeless : chr  "Yes" "Yes" NA "Yes" ...
##  $ Year         : int  2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
##  $ HH3          : int  0 1 NA 1 1 0 1 0 1 0 ...
##  $ HH4          : int  1 0 NA 1 0 0 1 0 1 1 ...
##  $ AD1          : int  0 1 NA 0 0 0 1 0 1 0 ...
##  $ AD1a         : int  NA 0 NA NA NA NA 1 NA 0 NA ...
##  $ AD2          : int  1 1 NA 0 0 0 1 0 1 0 ...
##  $ AD3          : int  1 1 NA 0 0 0 1 0 1 0 ...
##  $ index        : int  3 4 NA 2 1 0 6 0 5 1 ...
##  $ Gender       : chr  NA NA NA NA ...
```

```
##  $ USDAcat      : chr   "Low FS" "Low FS" NA "Low FS" ...
##  $ YR_2019      : chr   "Yes" "Yes" "Yes" "Yes" ...
##  $ YR_2020      : chr   "No" "No" "No" "No" ...
##  $ YR_2021      : chr   "No" "No" "No" "No" ...
##  $ YR_2022      : chr   "No" "No" "No" "No" ...
##  $ Ethn_Hisp    : chr   "No" "No" "Yes" "Yes" ...
##  $ Ethn_AIAN    : chr   "No" "No" "No" "No" ...
##  $ Ethn_Asian   : chr   "No" "No" "No" "No" ...
##  $ Ethn_Black   : chr   "Yes" "Yes" "No" "No" ...
##  $ Ethn_NHPI    : chr   "No" "No" "No" "No" ...
##  $ Ethn_White   : chr   "No" "No" "No" "No" ...
##  $ Ethn_Other   : chr   "No" "No" "No" "No" ...
##  $ Coll_BSN     : chr   "No" "Yes" "No" "No" ...
##  $ Coll_EDU     : chr   "No" "No" "No" "No" ...
##  $ Coll_ENGN    : chr   "No" "No" "Yes" "No" ...
##  $ Coll_HS      : chr   "No" "No" "No" "No" ...
##  $ Coll_LART    : chr   "Yes" "No" "No" "No" ...
##  $ Coll_SCI     : chr   "No" "No" "No" "No" ...
##  $ Coll_NRS     : chr   "No" "No" "No" "Yes" ...
##  $ Coll_PHR     : chr   "No" "No" "No" "No" ...
##  $ Coll_Other   : chr   "No" "No" "No" "No" ...
##  $ Coll_NA      : chr   "No" "No" "No" "No" ...
##  $ Gend_F       : chr   "No" "No" "No" "No" ...
##  $ Gend_M       : chr   "No" "No" "No" "No" ...
##  $ Gend_T       : chr   "No" "No" "No" "No" ...
##  $ Gend_NC      : chr   "No" "No" "No" "No" ...
##  $ Gend_Other   : chr   "No" "No" "No" "No" ...
##  $ Gend_PNA     : chr   "No" "No" "No" "No" ...
##  $ Gendercats   : chr   NA NA NA NA ...
```

```
str(extra)
```

```
## 'data.frame':    1743 obs. of  57 variables:
##  $ RespondentId
##  $ Q23..In.the.past.12.months.have.you.had.a.permanent.address.
##  $ Q27_1.Supplemental.Nutrition.Assistance.Program..SNAP..also.known.as.Food.StampsA.
##  $ Q27_2.Temporary.Assistance.for.Needy.Families..TANF.A.program..which.provides.fami
##  $ Q27_3.Women.Infant.and.Children..WIC.A.program.provides.foods..health.referrals..a
##  $ Q27_4.MedicaidA.health.insurance.program..which.provides.coverage.to.eligible.low.
##  $ Q27_5.Social.Security.Income..SSI.Monthly.benefits.to.people.with.limited.income.a
##  $ Q27_6.Military.Veteran.ServicesBenefits.such.as.financial.and.other.forms.of.assis
##  $ Q27_7.Unemployment.BenefitsFinancial.support.for.individuals.that.lost.their.job.a
##  $ Q27_8.None
##  $ Q27_9.Other
```

```
##  $ Q29_1.Free.emergency.food
##  $ Q29_2.Free.emergency.rent.assistance
##  $ Q29_3.Free.guidance.to.apply.for.government.food.programs..SNAP..WIC..etc.
##  $ Q29_4.Free.guidance.to.apply.for.government.housing.programs..rent.assistance..etc
##  $ Q29_5.Free.guidance.to.apply.for.other.government.programs..TANF..etc.
##  $ Q29_6.Dont.Know.Any
##  $ Q30_1.Free.emergency.food
##  $ Q30_2.Free.emergency.rent.assistance
##  $ Q30_3.Free.guidance.to.apply.for.government.food.programs..SNAP..WIC..etc.
##  $ Q30_4.Free.guidance.to.apply.for.government.housing.programs..rent.assistance..etc
##  $ Q30_5.Free.guidance.to.apply.for.other.government.programs..TANF..etc.
##  $ Q30_6.None
##  $ Q31..For.UTEP.students..how.difficult.is.it.to.get.emergency.food.assistance.
##  $ Q32_1.Lack.of.information..it.s.not.clear.which.sites.offer.emergency.food.assista
##  $ Q32_2.Lack.of.transportation
##  $ Q32_3.The.sites.are.located.too.far.away
##  $ Q32_4.The.sites.have.inconvenient.hours.of.operation
##  $ Q32_5.It.is.too.uncomfortable.to.pick.up.emergency.food.assistance
##  $ Q32_6.None
##  $ Q34_1.UTEP.Food.Pantry
##  $ Q34_2.Kelly.Center.for.Hunger.Relief
##  $ Q34_3.El.Pasoans.Fighting.Hunger.Food.Bank
##  $ Q34_4.Another.food.pantry.or.food.bank
##  $ Q41..How.often.have.you.found.it.difficult.to.concentrate.on.your.schoolwork.becau
##  $ Q42..Have.you.had.to.delay.the.completion.of.your.degree..because.you.did.not.have
##  $ Q43..How.often.do.you.think.about.delaying.the.completion.of.your.degree..because.
##  $ Q46..In.general..how.would.you.rate.your.mental.health..including.your.mood.and.yo
##  $ Q48..During.the.past.12.months..how.often.did.you.do.each.of.the.following..Mark.o
##  $ Q49..During.the.past.12.months..how.often.did.you.do.each.of.the.following..Mark.o
##  $ Q50..During.the.past.12.months..how.often.did.you.do.each.of.the.following..Mark.o
##  $ Q51..During.the.past.12.months..how.often.did.you.do.each.of.the.following..Mark.o
##  $ PermAdd
##  $ GovAss
##  $ AvailRes
##  $ MoreAvailRes
##  $ DiffLvlFA
##  $ ChallFA
##  $ EmergFood
##  $ DiffConcentrate
##  $ DelayComplDegree
##  $ TimeDelayComplDegree
##  $ RateMentalHealth
##  $ AttendProtest
##  $ AttendLocalComm
##  $ PoliticMess
```

```
##  $ CompValues
```

1.3   Filtering Out the 2022 Information from the Master Data Set

```
mas_22 <- master %>%
  filter(Year == "2022")%>%
  select(RespondentId, Age, Ethnicity, Income,
    Classification,FedAid,index,Gender,USDAcat)

head(mas_22)
```

```
##    RespondentId Age Ethnicity              Income Classification       FedAid index
## 1      95772100  NA         1  Less than $10,000       Doctoral        Loans     6
## 2      95772104   2         1 $30,000 to $39,999         Senior Work-study    NA
## 3      95772110  NA         1 $90,000 to $99,999         Junior        Loans    NA
## 4      95772115  NA         1 $20,000 to $29,999         Senior Work-study    NA
## 5      95772121  NA         1 $30,000 to $39,999         Senior        Loans     6
## 6      95772123   2         1   $100,000 or more         Junior Work-study    NA
##   Gender USDAcat
## 1      1    <NA>
## 2      1    <NA>
## 3      1    <NA>
## 4      1    <NA>
## 5      2    <NA>
## 6      2    <NA>
```

1.4   Merging the Filtered Info From the Master with the Extra

```
new_dat <- inner_join(extra, mas_22)
#head(new_dat)
dim(new_dat)
```

```
## [1] 1743    65
```

```
#names(new_dat)
```

1.5   Recoding Some Variables

```r
new_dat$USDAcat <-
  ifelse((new_dat$index==0|new_dat$index==1),"High FS",
  ifelse((new_dat$index==2|new_dat$index==3|new_dat$index==4),"Low FS",
  ifelse((new_dat$index==5|new_dat$index==6),"Very Low FS",
  new_dat$index)))

new_dat$FedAid <- ifelse((new_dat$FedAid=="Emergency Loan"),"Loans",
  ifelse((new_dat$FedAid=="UTEP's COVID CARES Act Fund"),"Covid Fund",
    new_dat$FedAid))

new_dat$Gender <- ifelse((new_dat$Gender==1),"Female",
                    ifelse((new_dat$Gender==2),"Male",
                     ifelse((new_dat$Gender==3),"Transgender","Others")))

new_dat$DelayComplDegree <- ifelse((new_dat$DelayComplDegree=="Yes,
                                    by 1 semester"), "1 Sem Lag",
  ifelse((new_dat$DelayComplDegree=="Yes, by 2 semesters or more"),
                                    "2/+ Sems Lag",
  ifelse((new_dat$DelayComplDegree=="No"), "None", new_dat$DelayComplDegree)))

names(new_dat)[names(new_dat) == "DelayComplDegree"] <- "Delay"

unique(new_dat$FedAid)
```

```
## [1] "Loans"       "Work-study"  "Scholarship" "Covid Fund"  "Other"
```

```r
unique(new_dat$USDAcat)
```

```
## [1] "Very Low FS" NA            "Low FS"      "High FS"
```

```r
unique(new_dat$Delay)
```

```
## [1] "2/+ Sems Lag"        "None"                "Yes, by 1 semester"
## [4] NA
```

2 Preliminary Visualizations of Key Analysis Variables

```r
vartab1 <- as.data.frame(table(new_dat$FedAid))
#prop <- paste0(round(vartab1[,2]/sum(vartab1[,2])*100,2),"%")
```

```
vartab2 <- as.data.frame(table(new_dat$USDACat))

vartab3 <- as.data.frame(table(new_dat$Delay))


# Bar plots of Federal Aid and Academic Delay

par(mfrow=c(2,1), mar=rep(3,4))

barplot(vartab1$Freq[order(vartab1$Freq,decreasing = T)],
        names = vartab1$Var1,col="blue",las=1,
         main = "Barplot of Federal Aid", cex.axis=0.5)

barplot(vartab3$Freq[order(vartab3$Freq,decreasing = T)],
        names = vartab3$Var1,
        col="green",
        main = "Barplot of Academic Delay", cex.axis=0.5)
```
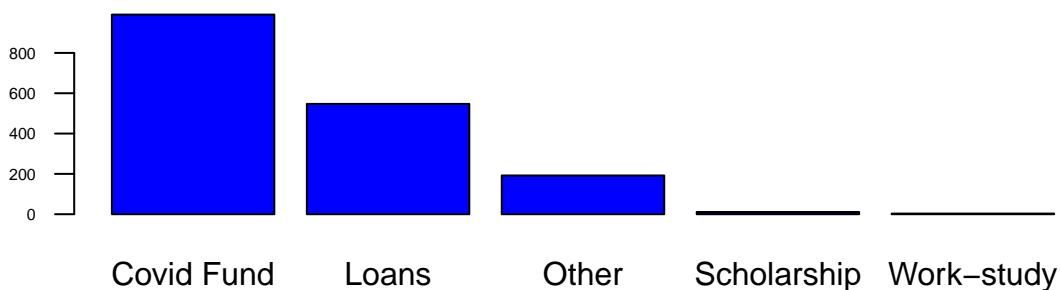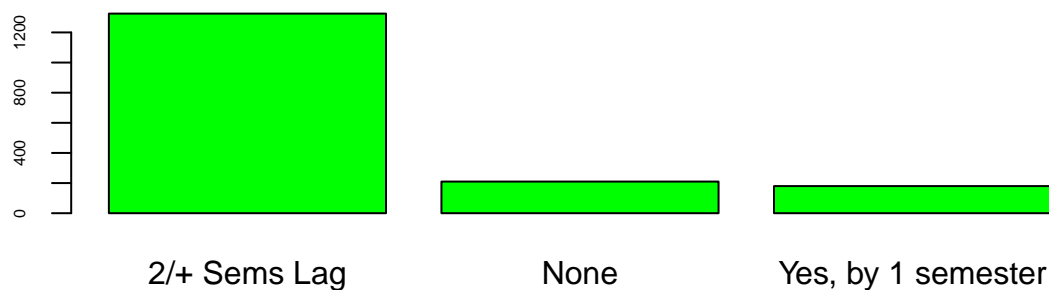
Barplot of Federal Aid

Barplot of Academic Delay

Comments: Viewing the above plots, we see that a chunk amount of Federal Aid came from the UTEP Covid Fund, followed by Loans, and others. As regards, the number of semesters students academic progress was delayed by, we see that the majority of students were delayed

for 1 semester for one reason or the other.

3   Visualizing Association Between Federal Aid and Food Security

3.1   Segmented Bar Chart of Federal Aid and Food Security

```r
# Categorical Bubble Plot
a <- ggplot(new_dat) +
    geom_count(aes(FedAid, USDAcat), color = "magenta") +
    labs(title = "Food Security By Federal Aid") +
    theme(plot.title = element_text(hjust = 0.5))



# segmented bar plot

seg <- new_dat %>%
  mutate(FedAid = FedAid,
         USDAcat = USDAcat) %>%
  group_by(FedAid, USDAcat) %>%
  summarise(freq = n()) %>%
  mutate(pct_lbl =(round(freq/sum(freq), 2)),
         lbl = scales::percent(pct_lbl))

head(seg)
```

```
## # A tibble: 6 x 5
## # Groups:   FedAid [2]
##   FedAid      USDAcat       freq pct_lbl lbl
##   <chr>       <chr>        <int>   <dbl> <chr>
## 1 Covid Fund  Low FS           1    0.09 9%
## 2 Covid Fund  Very Low FS      7    0.64 64%
## 3 Covid Fund  <NA>             3    0.27 27%
## 4 Loans       High FS          7    0.01 1%
## 5 Loans       Low FS          61    0.11 11%
## 6 Loans       Very Low FS    196    0.36 36%
```
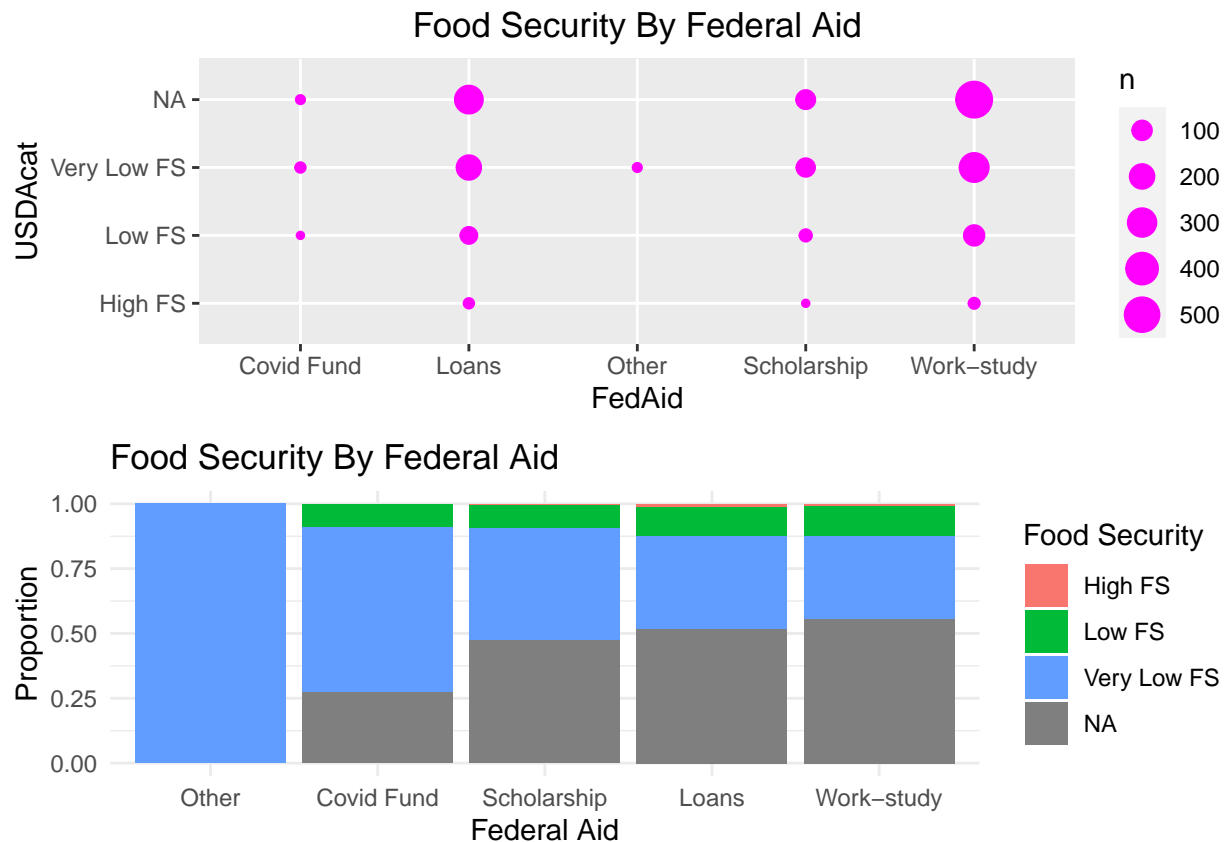
```r
b <- ggplot(seg, aes(reorder(FedAid,freq), freq, fill = USDAcat)) +
    geom_bar(stat = "identity", position = "fill") +
    ylab("Proportion") +
    xlab("Federal Aid") +
```

```
      ggtitle("Food Security By Federal Aid") +
      theme_minimal() +  labs(fill = "Food Security")
    #theme_minimal()

grid.arrange(a, b, nrow=2)
```


Comments: Seeing from the plots above, we see an interesting relationship between Federal Aid and Food security. In general, students who had high food security received less federal aid. Those who had low food security had a bit more federal aid than those who had high security. However, those who had very low food security had more federal aid, in general, than those who had low and high food security. Interestingly, only students in the very low food security category received "Other" federal aid, other than the remaining types of federal aid.

4   Visualizing Association Between Food Security and Degree Progress

5   creating stack bar plot

```r
stack <- new_dat %>%
  mutate(USDAcat = USDAcat,
         Delay = Delay) %>%
  group_by(USDAcat, Delay) %>%
  summarise(freq = n())%>%
  mutate(pct = freq/sum(freq),
         lbl = scales::percent(pct))
  head(stack)
```

```
## # A tibble: 6 x 5
## # Groups:   USDAcat [2]
##    USDAcat Delay                 freq      pct lbl
##    <chr>   <chr>                <int>   <dbl> <chr>
## 1 High FS 2/+ Sems Lag             1 0.0556  6%
## 2 High FS None                    17 0.944   94%
## 3 Low FS  2/+ Sems Lag            20 0.103   10.3%
## 4 Low FS  None                   149 0.768   76.8%
## 5 Low FS  Yes, by 1 semester     24 0.124   12.4%
## 6 Low FS  <NA>                     1 0.00515 0.5%
```

```r
p <- ggplot(stack, aes(fill=reorder(Delay, freq), y=freq,
  x=USDAcat, label=Delay)) +
  geom_bar(position="fill", stat="identity") +
  geom_text(aes(label =lbl), position = position_fill(vjust=
      0.5), size=2) +
  ylab("Proportion") +
  xlab("Food Security") +
  ggtitle("Association between Academic Delay
    and Food Security") +
  theme_minimal() +  labs(fill = "Delay")



tab <- new_dat %>%
  group_by(USDAcat,index, Gender,Delay) %>%
  summarise(freq = n()) %>%
```

```
  mutate(prop = round(freq*100/sum(freq),2))
head(tab)
```

```
## # A tibble: 6 x 6
## # Groups:   USDAcat, index, Gender [5]
##    USDAcat index Gender Delay          freq  prop
##    <chr>   <int> <chr>  <chr>         <int> <dbl>
## 1 High FS      0 Female None              3   100
## 2 High FS      0 Male   None              1   100
## 3 High FS      1 Female None             10   100
## 4 High FS      1 Male   2/+ Sems Lag      1    25
## 5 High FS      1 Male   None              3    75
## 6 Low FS       2 Female 2/+ Sems Lag      3    15
```

```r
# creating a flow chart
f <- ggplot(tab,
       aes(axis1 = USDAcat,
           axis2 = Delay,
           y = freq)) +
  geom_alluvium(aes(fill=USDAcat)) +
  geom_stratum() +
  geom_text(stat = "stratum",
          aes(label = after_stat(stratum))) +
  scale_x_discrete(limits = c("Food Insecurity", "Degree
    Progress"), expand = c(.1, .1)) +
  labs(title = "UTEP 2022 Data", subtitle = "Stratified by
    Academic Delay and Food Security", y = "Frequency") +
  theme_minimal()

grid.arrange(p, f, nrow=2)
```
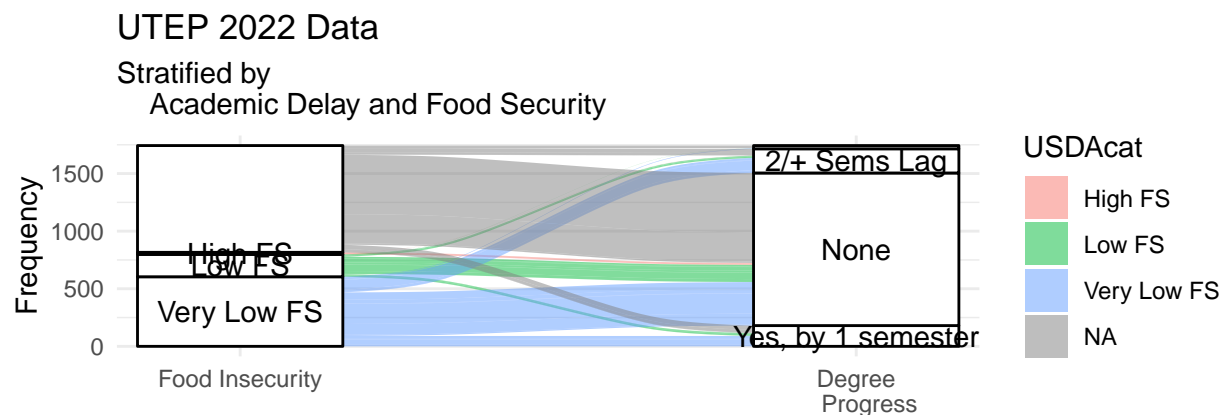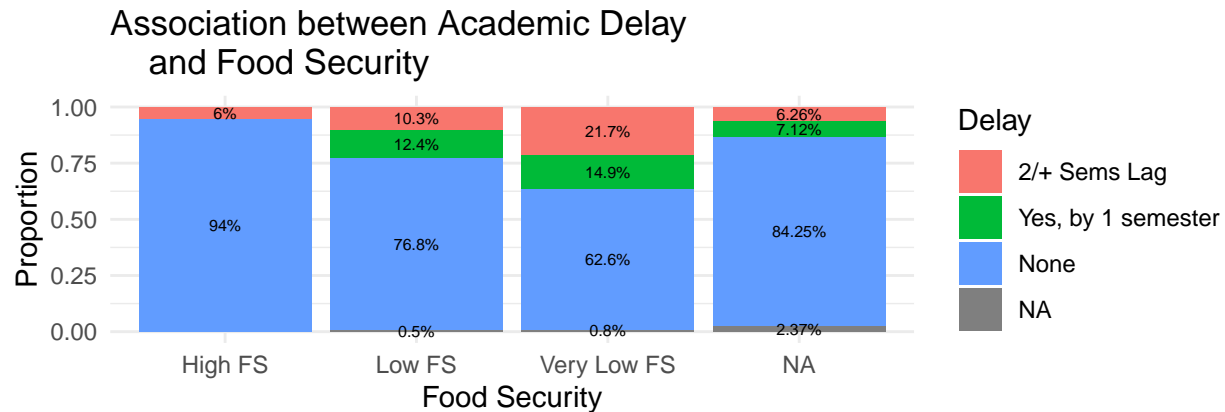
Comment: The two plots (the stacked bar graph on top and the flow chart) above reveal an implicitly inverse relationship between UTEP food security and the items pertaining to concentration on school or degree progress. It can be observed that, students who had high food security had maximum to full concentration on their studies and had no semester delays in their degree progress. Students who had low food security had at least one semester delay in their academic progress, whereas students who had very low food security, had at least two semester delays in their degree progress since they lost full concentration on their studies. Thus, it can be concluded that food insecurity had a negative effect on students concentration on studies.

6   Visualizing the Association Between Gender and Degree Progress

```
par(mar=rep(1,4))

r <- ggplot(tab, mapping = aes(x=Gender, y=Delay,
  fill=prop)) +
  geom_tile(color = "white", linewidth = 1.5, linetype=1) +
  # coord_fixed(ratio = 0.) +
```

```r
  labs(title = "Heatmap:School Delay By Gender")

s <- ggplot(new_dat) +
  geom_count(aes(Delay, Gender),color="blue") +
  labs(title = "Bubble Plot:School Delay By Gender") +
  theme(plot.title = element_text(hjust = 0.5))

theme(legend.position = "none",
    axis.text.y = element_text(size=10), # adjust the size of the y-axis text

    axis.title.y = element_text(size=12), # adjust the size of the y-axis title

    axis.text.x = element_text(size=10), # adjust the size of the x-axis text

    axis.title.x = element_text(size=12), # adjust the size of the x-axis title

    plot.title = element_text(size=8), # adjust the size of the plot title

    panel.grid.major.y = element_blank(), # remove the horizontal gridlines

    panel.grid.minor.y = element_blank(), # remove the horizontal gridlines

    plot.margin = unit(c(1, 0.5, 0.5, 3), "lines"))
```

```
## List of 9
##  $ axis.title.x      :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : num 12
##   ..$ hjust       : NULL
##   ..$ vjust       : NULL
##   ..$ angle       : NULL
##   ..$ lineheight  : NULL
##   ..$ margin      : NULL
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.title.y      :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : num 12
##   ..$ hjust       : NULL
```

```
##    ..$ vjust       : NULL
##    ..$ angle       : NULL
##    ..$ lineheight  : NULL
##    ..$ margin      : NULL
##    ..$ debug       : NULL
##    ..$ inherit.blank: logi FALSE
##    ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.x      :List of 11
##    ..$ family      : NULL
##    ..$ face        : NULL
##    ..$ colour      : NULL
##    ..$ size        : num 10
##    ..$ hjust       : NULL
##    ..$ vjust       : NULL
##    ..$ angle       : NULL
##    ..$ lineheight  : NULL
##    ..$ margin      : NULL
##    ..$ debug       : NULL
##    ..$ inherit.blank: logi FALSE
##    ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.y      :List of 11
##    ..$ family      : NULL
##    ..$ face        : NULL
##    ..$ colour      : NULL
##    ..$ size        : num 10
##    ..$ hjust       : NULL
##    ..$ vjust       : NULL
##    ..$ angle       : NULL
##    ..$ lineheight  : NULL
##    ..$ margin      : NULL
##    ..$ debug       : NULL
##    ..$ inherit.blank: logi FALSE
##    ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ legend.position  : chr "none"
##  $ panel.grid.major.y: list()
##    ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ panel.grid.minor.y: list()
##    ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ plot.title       :List of 11
##    ..$ family      : NULL
##    ..$ face        : NULL
##    ..$ colour      : NULL
##    ..$ size        : num 8
##    ..$ hjust       : NULL
##    ..$ vjust       : NULL
```

```
##    ..$ angle      : NULL
##    ..$ lineheight : NULL
##    ..$ margin     : NULL
##    ..$ debug      : NULL
##    ..$ inherit.blank: logi FALSE
##    ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ plot.margin    : 'simpleUnit' num [1:4] 1lines 0.5lines 0.5lines 3lines
##    ..- attr(*, "unit")= int 3
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

```
ggpubr::ggarrange(r, s, nrow = 2)
```
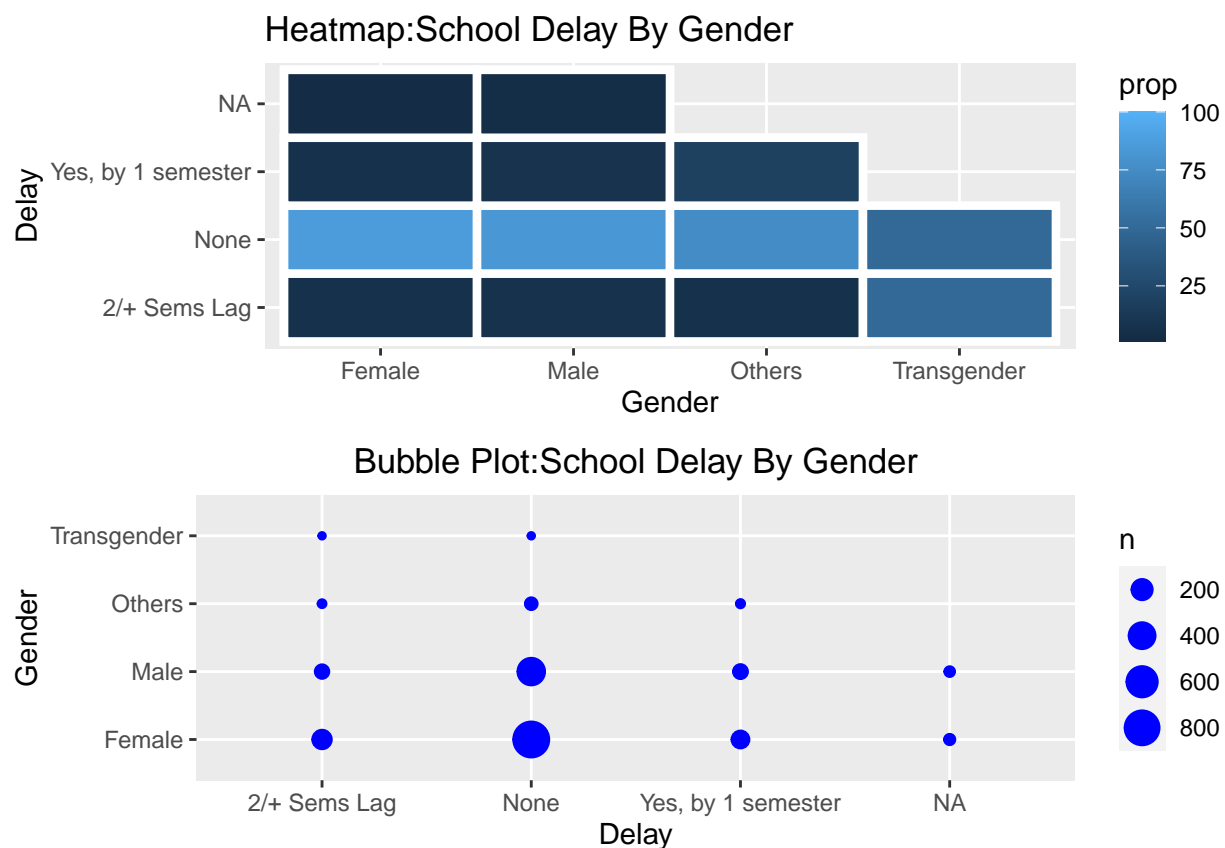

Comments: From the Heat map and Categorical bubble plot above, we can generally infer that, for those who had one or two semesters delay in their degree progress, there were more females than males. And it is the same case that when we observe among those in the "None" Delay category, the females were more than the males. This could be due to the fact that more respondents declared their gender as "female" than those who declared as "male", and thus resulted in the data collecting more information on the female gender than on the male gender.