

# STAT 527 HW 4

Satoshi Ido (ID: 34788706)

13 February 2023

pre-setting

```
library("ggplot2")  
# change the working directory for knitting the document  
setwd("/Users/satoshiido/Documents/statistical-analysis/Purdue_STAT527")
```

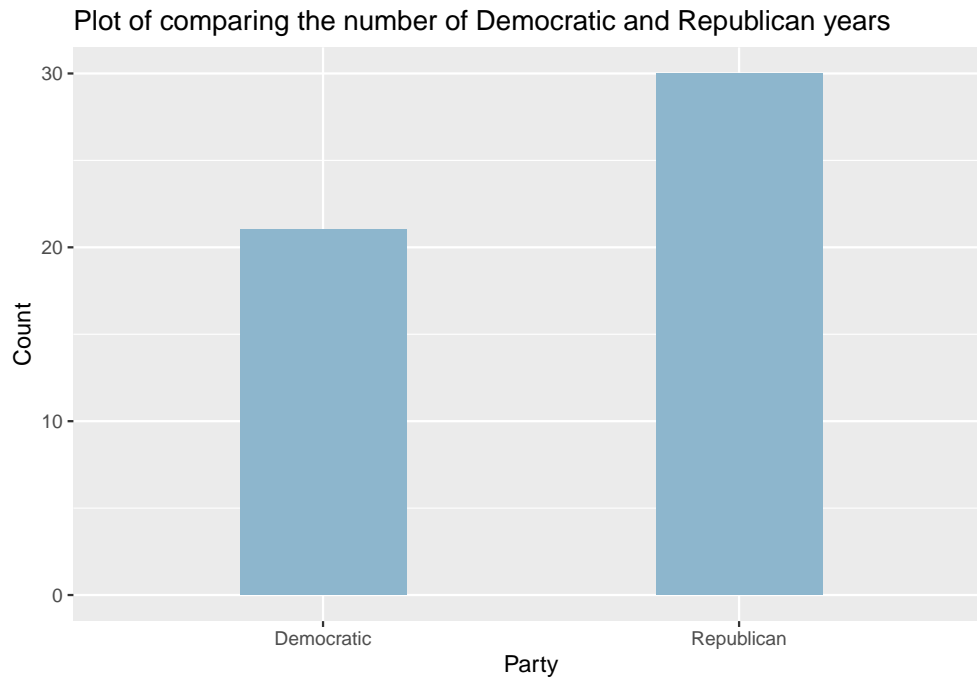
1. (12 points) US economics data (Table 4.1) in Chapter 4.

```
# import data  
df <- read.delim("economydata.txt", header = TRUE, sep = ",", dec = ".")  
head(df)
```

##	YEAR	UNEMPLOYMENT	GDPGROWTH	INFLATION	SURPLUS	PARTY
## 1	1960	5.5	2.5	1.4	0.1	R
## 2	1961	6.7	2.3	0.7	-0.6	R
## 3	1962	5.5	6.1	1.3	-1.3	D
## 4	1963	5.7	4.4	1.6	-0.8	D
## 5	1964	5.2	5.8	1.0	-0.9	D
## 6	1965	4.5	6.4	1.9	-0.2	D

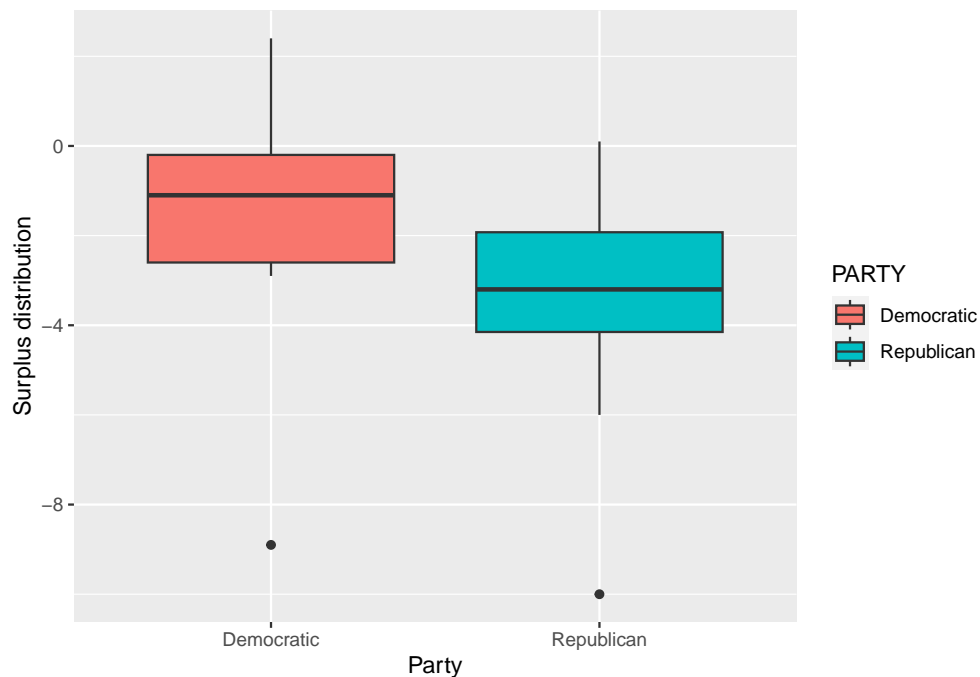
- (a) Obtain the bar graph comparing the number of Democratic and Republican presidential budget years. Be sure to relabel everything in the graphs appropriately.

```
# rename the value  
df$PARTY[df$PARTY == "R"] <- "Republican"  
df$PARTY[df$PARTY == "D"] <- "Democratic"  
  
# plot  
ggplot(df, aes(x = PARTY)) +  
  geom_bar(width = 0.4, fill = "lightskyblue3") +  
  labs(x = "Party", y = "Count") +  
  ggtitle("Plot of comparing the number of Democratic and Republican years")
```



- (b) Obtain the side-by-side boxplot comparing Democratic and Republican presidential budget years, for the variable surplus. Be sure to relabel everything in the graphs appropriately.

```
ggplot(df, aes(x = PARTY, y = SURPLUS, fill = PARTY)) +
  geom_boxplot() +
  labs(x = "Party", y = "Surplus distribution")
```



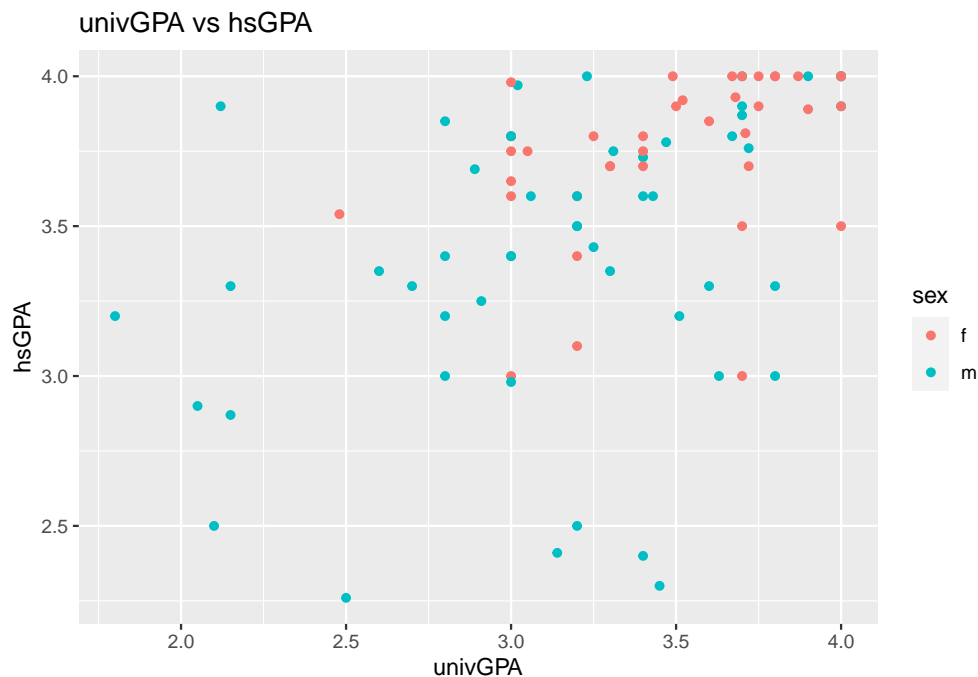
(18 points) GPA data (Table 5.1) in Chapter 5.

```
# import data
df2 <- read.delim("GPAdata.txt", header = TRUE, sep = "", dec = ".")
head(df2)
```

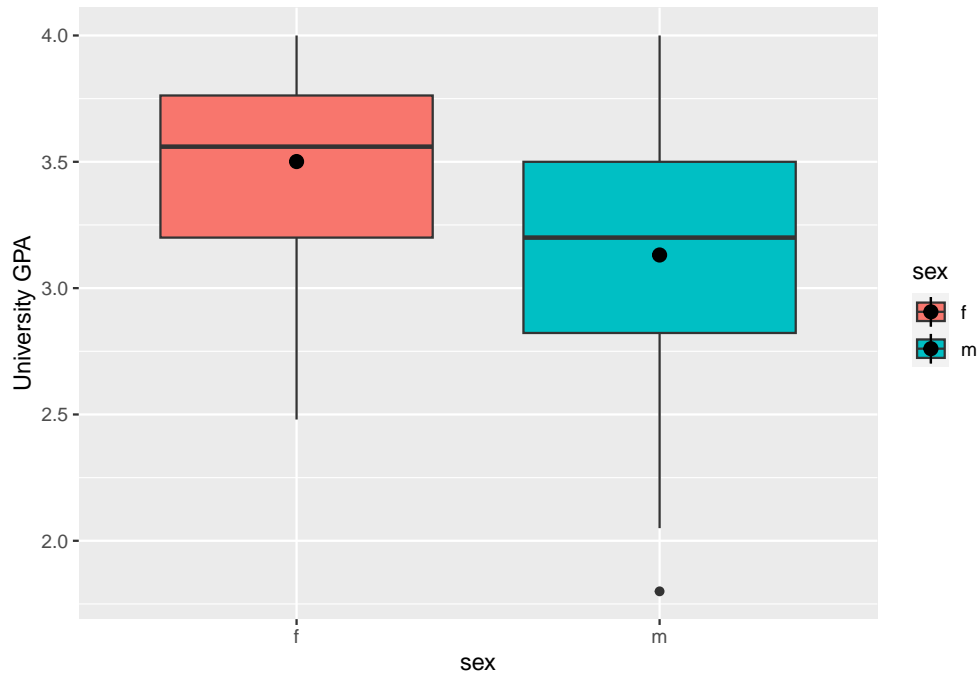
```
##   univGPA ACT hsGPA sex housing
## 1    3.40  24  3.73   m      o
## 2    3.25  30  3.43   m      o
## 3    3.47  24  3.78   m      o
## 4    3.63  24  3.00   m      o
## 5    1.80  27  3.20   m      o
## 6    3.60  19  3.30   m      r
```

- (a) Graphically explore the college GPA data for potential associations of univGPA with other variables besides ACT.

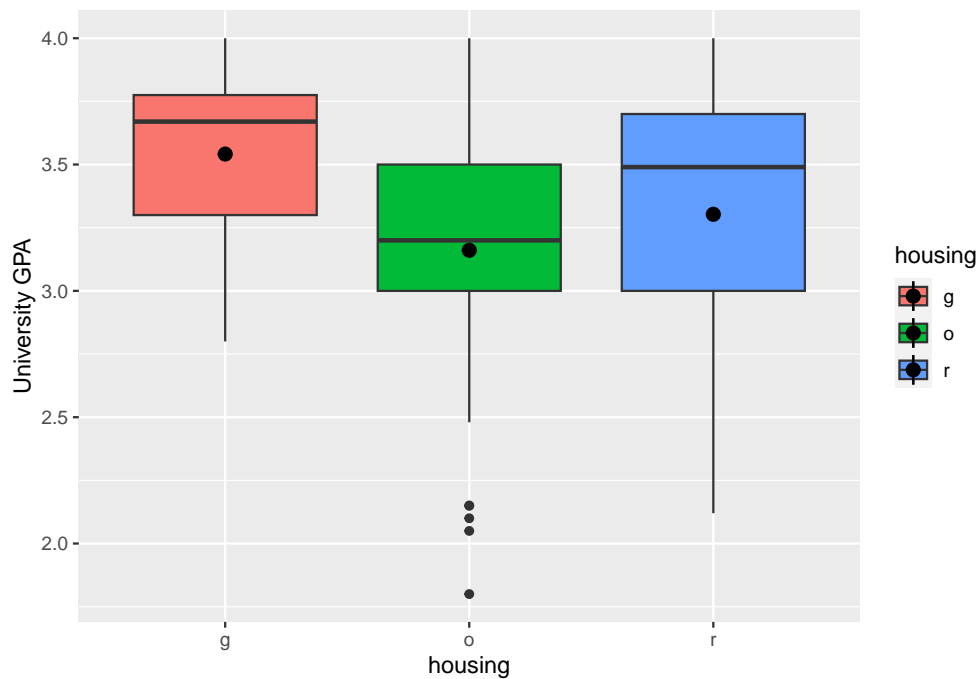
```
# check the correlation between univGPA and hsGPA
## It seems there is no specific correlation
## if comparing only between university GPA and highschool GPA
## Yet, by grouping up with sex, we can a trend that is, female is better
ggplot(df2, aes(x = univGPA, y = hsGPA, color = sex)) +
  geom_point() +
  ggtitle("univGPA vs hsGPA")
```



```
# check the boxplot between univGPA and sex, housing
## Female students seem to have better university GPA scores.
ggplot(df2, aes(x = sex, y = univGPA, fill = sex)) +
  geom_boxplot() +
  labs(x = "sex", y = "University GPA") +
  stat_summary(fun.y = "mean")
```



```
## Students living at housing "g" seem to have better university GPA scores.
ggplot(df2, aes(x = housing, y = univGPA, fill = housing)) +
  geom_boxplot() +
  labs(x = "housing", y = "University GPA") +
  stat_summary(fun.y = "mean")
```

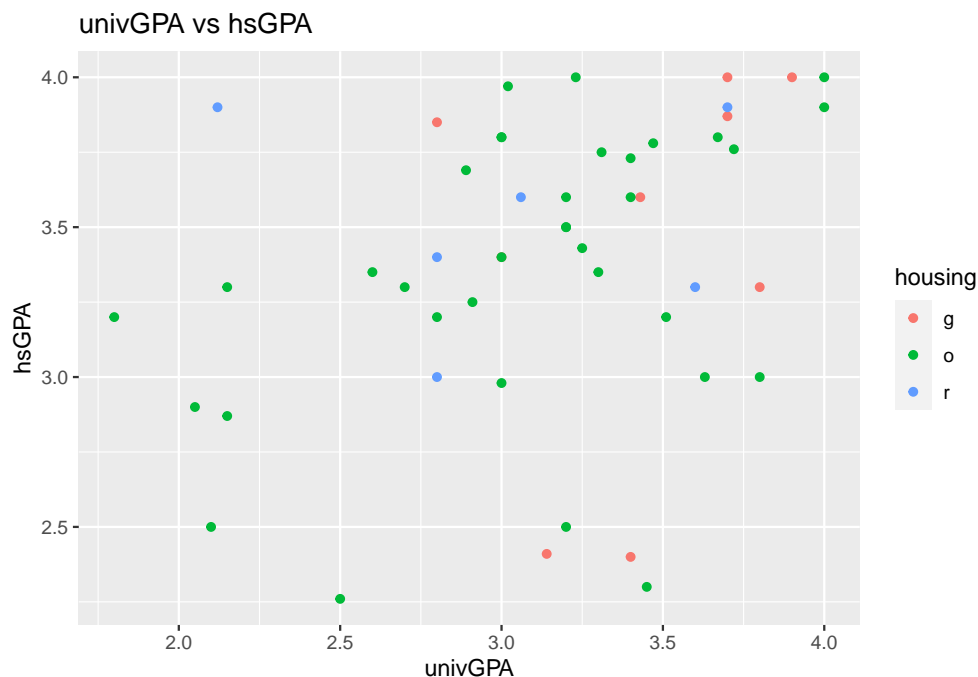


- (b) Separate out the males-only observations in the college GPA data. Graphically explore the males-only data for potential associations of univGPA with other variables besides ACT.

```
# extract male data
df2_m <- df2[df2$sex == "m", ]
head(df2_m)
```

```
##   univGPA ACT hsGPA sex housing
## 1    3.40  24  3.73   m      o
## 2    3.25  30  3.43   m      o
## 3    3.47  24  3.78   m      o
## 4    3.63  24  3.00   m      o
## 5    1.80  27  3.20   m      o
## 6    3.60  19  3.30   m      r
```

```
# check the correlation between univGPA and hsGPA
## Still there is no specific trend between high school GPA and university GPA
## Many male students live at housing "o".
## It's not good to compare housing as numbers of students lean on housing "o"
ggplot(df2_m, aes(x = univGPA, y = hsGPA, color = housing)) +
  geom_point() +
  ggtitle("univGPA vs hsGPA")
```



- (c) Separate out the females-only observations in the college GPA data. Graphically explore the females only data for potential associations of univGPA with ACT as well as other variables.

```
# extract female data
df2_f <- df2[df2$sex == "f", ]
head(df2_f)
```

```
##   univGPA ACT hsGPA sex housing
## 11    3.00  14  3.00   f      o
```

```
## 12    3.90  32  3.89  f    r
## 13    3.52  25  3.92  f    g
## 18    3.25  22  3.80  f    r
## 22    3.40  24  3.70  f    o
## 26    4.00  29  4.00  f    g
```

```
# check the correlation between univGPA and ACT
## there is a relatively small positive correlation between
## university GPA and ACT among female students
## Female students living at housing "g" seems to have better scores
ggplot(df2_f, aes(x = univGPA, y = ACT, color = housing)) +
  geom_point() +
  ggtitle("univGPA vs ACT")
```

