Lab 9

Alexander Hernandez

09/22/2022

1) data "vacation" provided in the link below describe a sample of 200 Chicago households regarding their vacation. The data includes the following variables: miles, income, age, kids

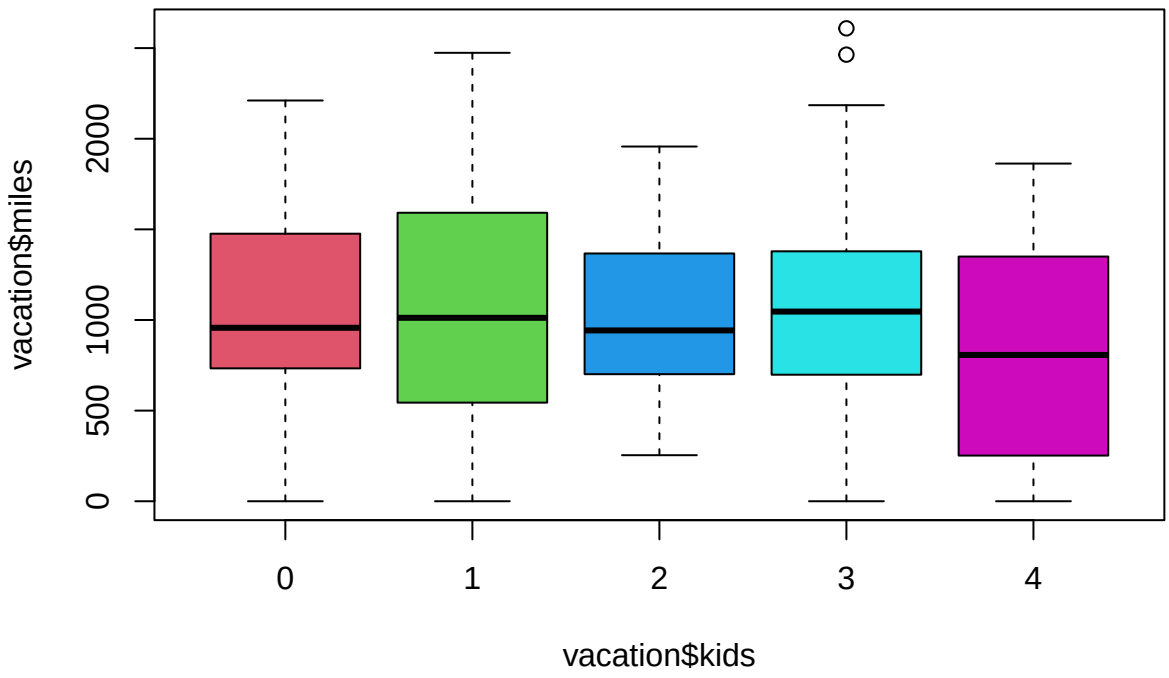
```
# http://www.principlesofeconometrics.com/poe4/data/stata/vacation.dta
```

a) a) Import the data in R

```
library(haven)
vacation = read_dta('C:\\repos\\STAT 50001\\Lab 9\\vacation.dta')
```

b) Display the miles distribution based on the number of kids by drawing parallel box-plot

```
boxplot(vacation$miles ~ vacation$kids, col=c(2,3,4,5,6))
```



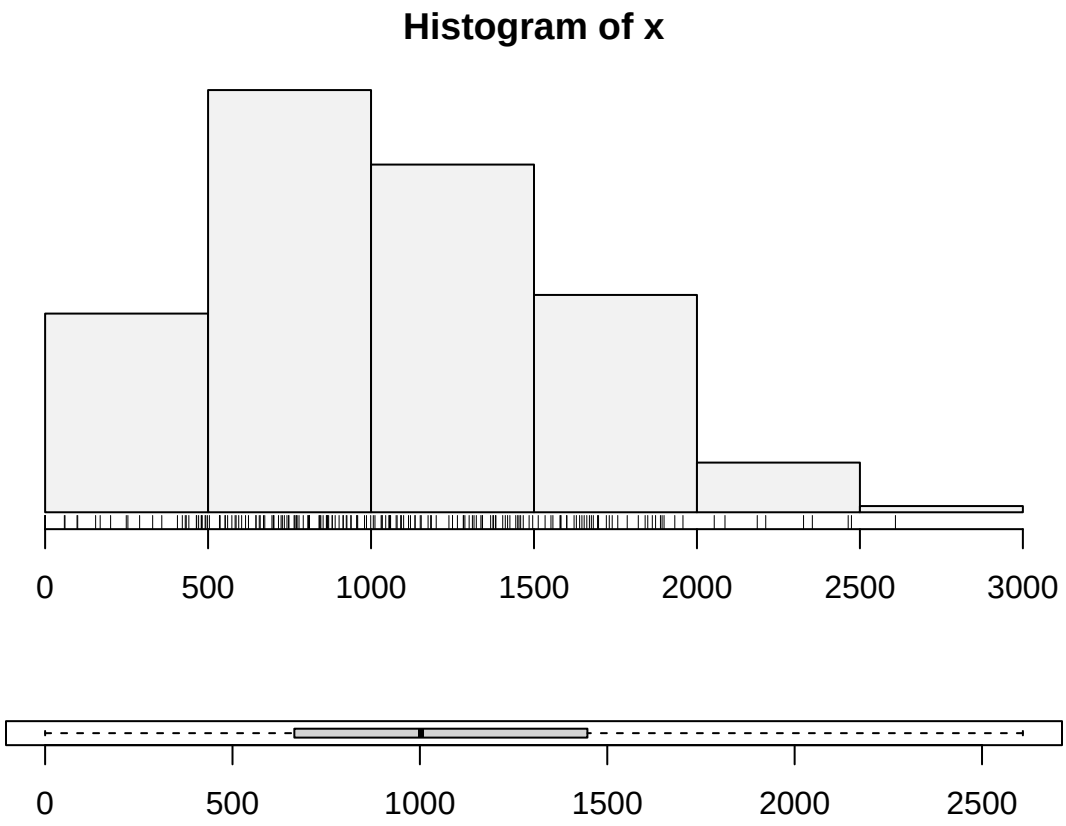
c) Draw histogram along with boxplot of the income data

```
library(UsingR)
```

```
## Loading required package: MASS
## Loading required package: HistData
## Loading required package: Hmisc
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##     format.pval, units
##
## Attaching package: 'UsingR'
## The following object is masked _by_ '.GlobalEnv':
##
##     vacation
```

```
## The following object is masked from 'package:survival':
##
##     cancer
```

```
simple.hist.and.boxplot(vacation$miles)
```



2) The following are the head circumferences (centimeters) at birth of 15 infants. Construct 95% CI for head circumferences (cm) at birth of all infants born at the local hospital

```
# 33.38 32.15 33.99 34.10 33.97 34.34 33.95 33.85 34.23 32.73 33.46 34.13 34.45 34.19 34.05
infants = scan('C:\\repos\\STAT 50001\\Lab 9\\data.txt')
t.test(infants, conf.level=0.95)$conf.int
```

```
## [1] 33.44895 34.14705
## attr(,"conf.level")
## [1] 0.95
```

3) Hurricane Data

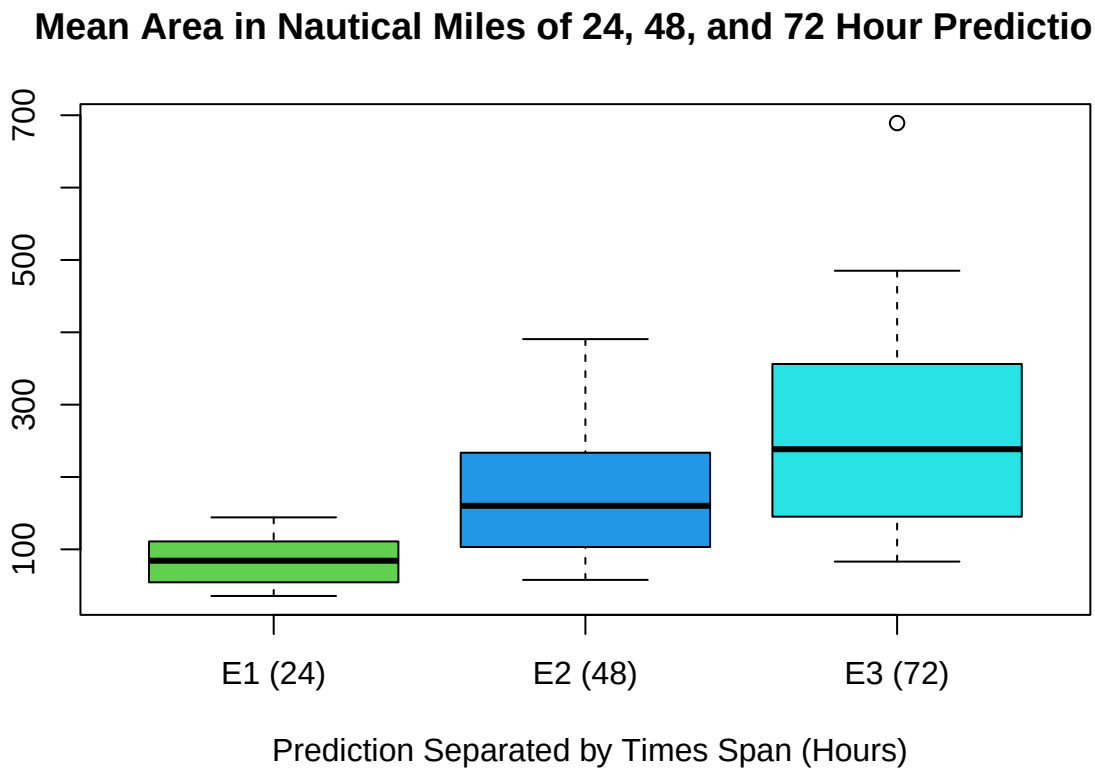
```
# https://dasl.datadescription.com/datafile/tracking-hurricanes-2016/
```

a) Import the data in R

```
hurricane = read.table('C:\\repos\\STAT 50001\\Lab 9\\tracking-hurricanes-2016.txt',
                       sep="\t", header=TRUE)
```

b) Display the 24-, 48- and 72-hours errors creating appropriate graph

```
names(hurricane) = c("Year", "E1", "E2", "E3")
boxplot(hurricane$E1, hurricane$E2, hurricane$E3,
        names=c("E1 (24)", "E2 (48)", "E3 (72)"), col=c(3,4,5),
        main="Mean Area in Nautical Miles of 24, 48, and 72 Hour Predictions",
        xlab=" Prediction Separated by Times Span (Hours)")
```



c) Construct 90% CI for 72 hours prediction errors

```
t.test(hurricane$E3, conf.level=0.9)$conf.int
```

```
## [1] 230.5047 294.0230
## attr(,"conf.level")
## [1] 0.9
```

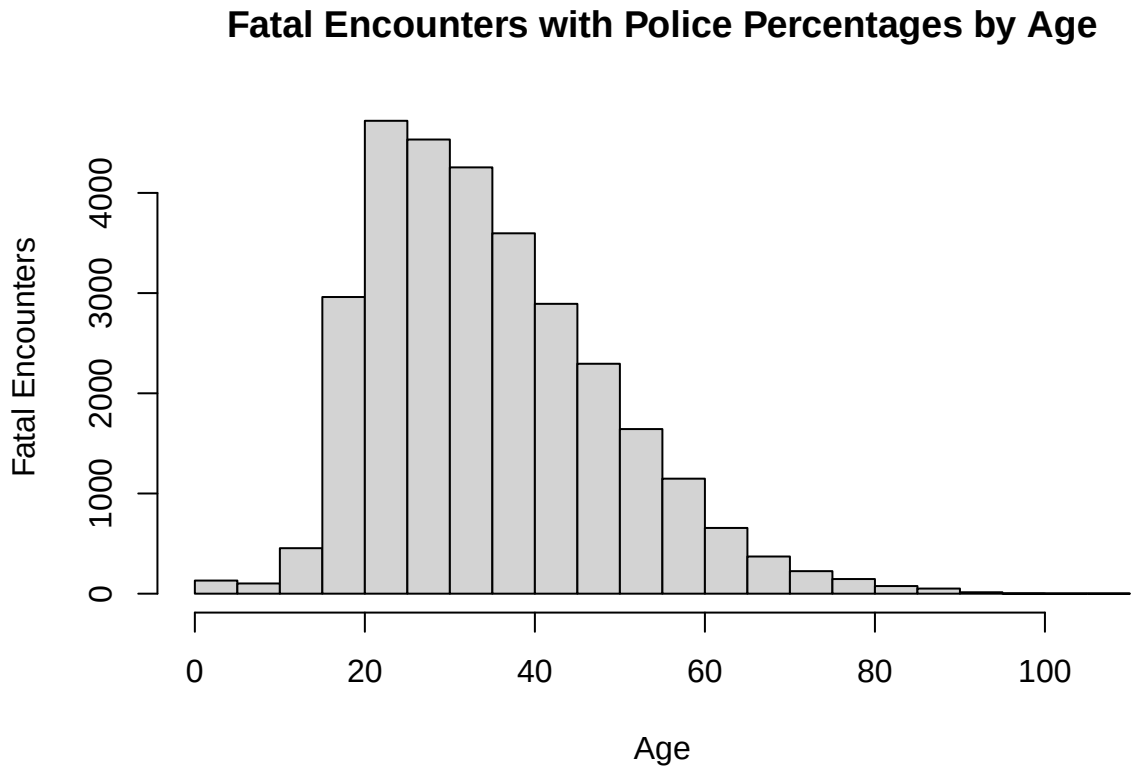
4) Fatal Encounters

```
# https://docs.google.com/spreadsheets/d/1dKmaV_JiWcG8XBoRgP8b4e9EopkpgtYFLYnyspvzAsE/edit#gid=0
```

a) Import the data in R and display age distribution using histogram

```
# Unable to get googlesheets4 working
library(readxl)

fatalities = read_excel("C:\\repos\\STAT 50001\\Lab 9\\FATAL ENCOUNTERS DOT ORG SPREADSHEET (See Read m
                        sheet=1,
                        range = cell_cols("A:G"))

hist(as.numeric(na.omit(fatalities$Age)),
     main = "Fatal Encounters with Police Percentages by Age",
     xlab = "Age",
     ylab = "Fatal Encounters")
```

```
## Warning in hist(as.numeric(na.omit(fatalities$Age)), main = "Fatal Encounters
## with Police Percentages by Age", : NAs introduced by coercion
```



b) Display the gender distribution using a pie chart

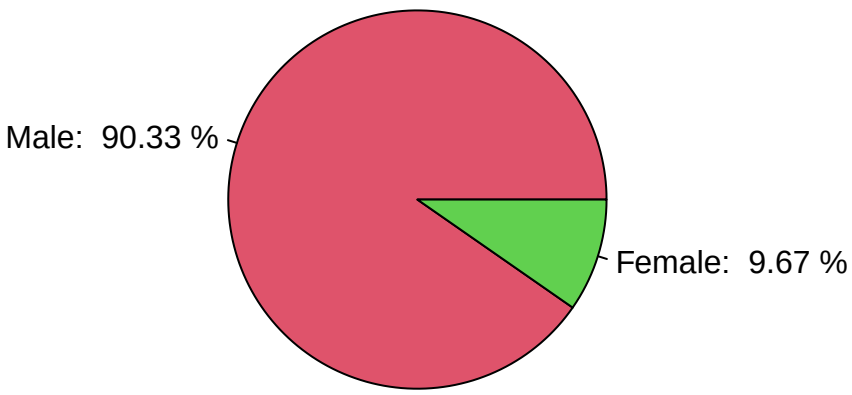
```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

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

```
## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
male = length(which(fatalities$Gender == "Male"))
female = length(which(fatalities$Gender == "Female"))
total = male + female

male_percent = round((male / total), digits=4) * 100
female_percent = round((female / total), digits=4) * 100

pie_labels = c(paste("Male: ", male_percent, "%"),
               paste("Female: ", female_percent, "%"))

pie(rbind(male, female),
    col=c(2,3),
    labels = pie_labels,
    main = "Fatal Encounters with Police Percentages by Gender")
```



c) Display race distribution using a pie chart

```
race_data = fatalities %>% count(fatalities$`Race with imputations`,
                                 sort = TRUE)
names(race_data) = c("Races", "Frequency")
race_data
```

```
## # A tibble: 11 x 2
##    Races                    Frequency
##    <chr>                        <int>
##  1 European-American/White      14729
##  2 African-American/Black        8545
##  3 Hispanic/Latino               5111
##  4 Race unspecified              1289
##  5 NA                             863
##  6 Asian/Pacific Islander         576
##  7 Native American/Alaskan        323
##  8 Middle Eastern                  53
##  9 <NA>                             6
## 10 HIspanic/Latino                  2
## 11 european-American/White          1
```

```
pie(race_data$Frequency,
    labels=(race_data$Races),
    main = "Fatal Encounters with Police Percentages by Race")
```