

Unit 6 – R for Graphs
Homework
SOLUTIONS

Download from the course website

[descriptive_gss.Rdata](#)

Initialize R Studio Session.

```
setwd("~/Desktop")      # Set working directory
getwd()                 # Check working directory

## [1] "/Users/cbigelow/Desktop"

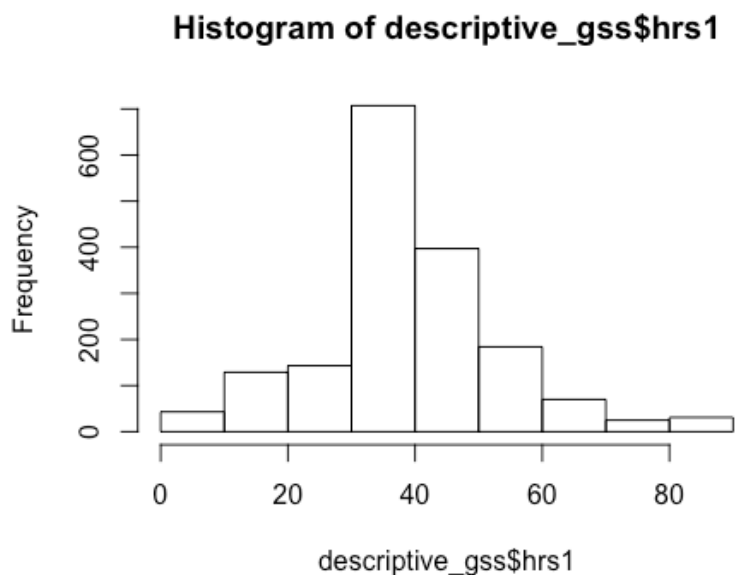
options(scipen=999)     # Turn off scientific notation
rm(list = ls())         # Clear the Decks
```

Input data

```
load(file="descriptive_gss.Rdata")
#str(descriptive_gss)      # Not knitting this
```

Q1 - “No frills” histogram of hrs1

```
hist(descriptive_gss$hrs1)
```



Q2 - Histogram of hrs1 with overlay normal and aesthetics

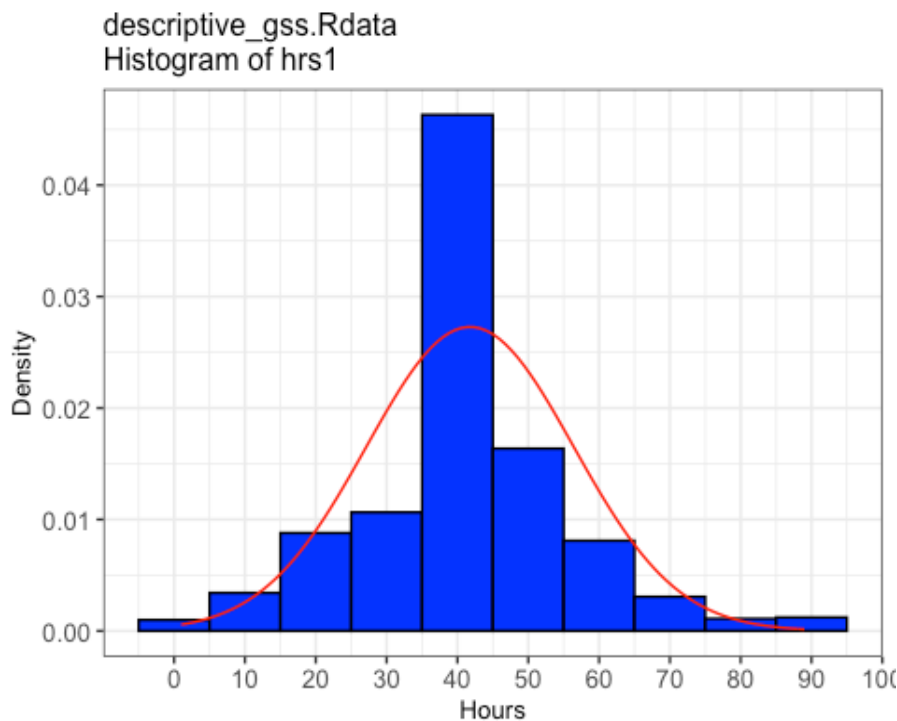
```
library(ggplot2)
library(stargazer)

# Preliminary - min and max for setting y-axis tick marks
descriptive_gss <- as.data.frame(descriptive_gss) # stargazer works on dataframes only
stargazer(descriptive_gss[c("hrs1")],type="text",summary.stat=c("min","max"))

##
## =====
## Statistic Min   Max
## -----
## hrs1      1.000 89.000
## -----

ggplot(data=descriptive_gss, aes(x=hrs1)) +
  geom_histogram(binwidth=10, colour="black", fill="blue", aes(y=..density..)) +
  stat_function(fun=dnorm, colour="red",
    args=list(mean=mean(descriptive_gss$hrs1,na.rm=TRUE),
    sd=sd(descriptive_gss$hrs1, na.rm=TRUE))) +
  scale_x_continuous(breaks=seq(0, 100, 10))+
  ggtitle("descriptive_gss.Rdata \nHistogram of hrs1") +
  xlab("Hours") +
  ylab("Density") +
  theme_bw() +
  theme(axis.text=element_text(size=10),
    axis.title=element_text(size=10),
    plot.title=element_text(size=12))

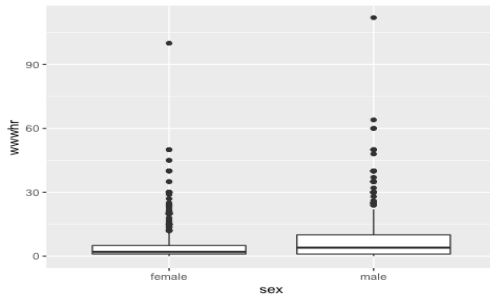
## Warning: Removed 1036 rows containing non-finite values (stat_bin).
```



Q3 - Boxplot summary of wwahr, separately by sex

```
library(ggplot2)
library(stargazer)

ggplot(data=descriptive_gss, aes(x=sex,y=wwahr)) +           # no frills
  geom_boxplot(na.rm=TRUE)
```

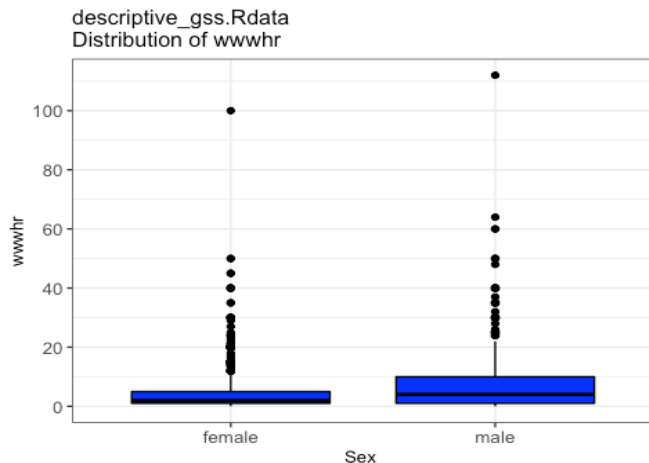


```
# Preliminary - min and max for setting y-axis tick marks
descriptive_gss <- as.data.frame(descriptive_gss)           # stargazer works on dataframes only
stargazer(descriptive_gss[c("wwahr")],type="text",summary.stat=c("min","max"))
```

```
##
## =====
## Statistic  Min    Max
## -----
## wwahr      0.000 112.000
## -----
```

```
ggplot(data=descriptive_gss, aes(x=factor(sex), y=wwahr)) +
  geom_boxplot(color="black", fill="blue") +
  ggtitle("descriptive_gss.Rdata\nDistribution of wwahr") +
  scale_y_continuous(breaks=seq(0, 120, 20))+
  xlab("Sex ") +
  ylab("wwahr") +
  theme(legend.position = "none") +
  theme_bw() +
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=10),
        plot.title=element_text(size=12))
```

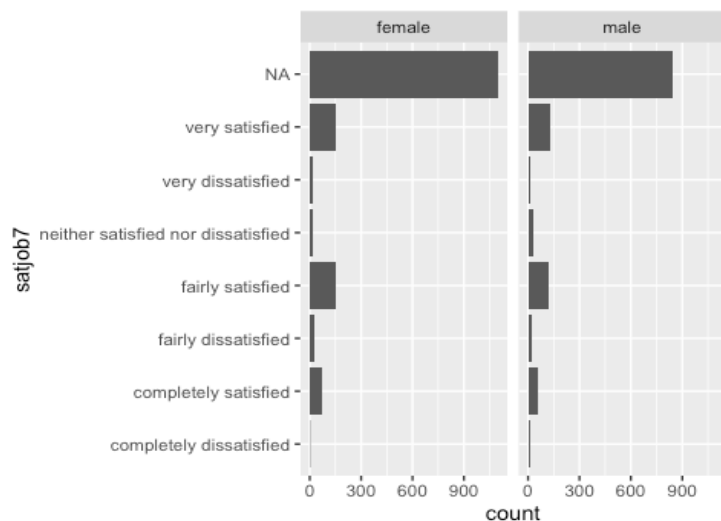
```
## Warning: Removed 1191 rows containing non-finite values (stat_boxplot).
```



Q4 - Bar chart summary of satjob7, separately by sex

```
library(ggplot2)
library(tidyverse)

ggplot(data=descriptive_gss,aes(x=satjob7))+      # no frills
  geom_bar(na.rm=T)+
  coord_flip() +                                  # flip the axes for easier readability!
  facet_wrap(~sex)
```



```
ggplot(data=descriptive_gss,aes(x=satjob7))+      # w aesthetics
  geom_bar(stat="count", width=0.7, fill="blue", color="black",na.rm=TRUE)+
  xlab(" ") +                                     # original x-axis
  ylab("Frequency") +                             # original y-axis
  coord_flip() +                                  # flip the axes for easier readability
  facet_wrap(~sex) +
  theme_minimal() +
  ggtitle("Job Satisfaction\nby Sex")
```



```
# Just for fun - Plot complete data ONLY
q4data <- descriptive_gss %>% select("satjob7","sex")
q4data <- na.omit(q4data)
ggplot(data=q4data,aes(x=satjob7))+
  geom_bar(stat="count", width=0.7, fill="blue", color="black",na.rm=TRUE)+
  xlab(" ") + # original x-axis
  ylab("Frequency") + # original y-axis
  coord_flip() + # flip the axes for easier readability
  facet_wrap(~sex) +
  theme_minimal() +
  ggtitle("Job Satisfaction by Sex\nComplete data ONLY")
```

