Homework 10

Danielle Banks, Amanda Norton, and Miles Tweed

2/23/2021

Problem 1

```
health<-as.tibble(healthcare_dataset_stroke_data)
```

```
## Warning: `as.tibble()` was deprecated in tibble 2.0.0.
## Please use `as_tibble()` instead.
## The signature and semantics have changed, see `?as_tibble`.
```

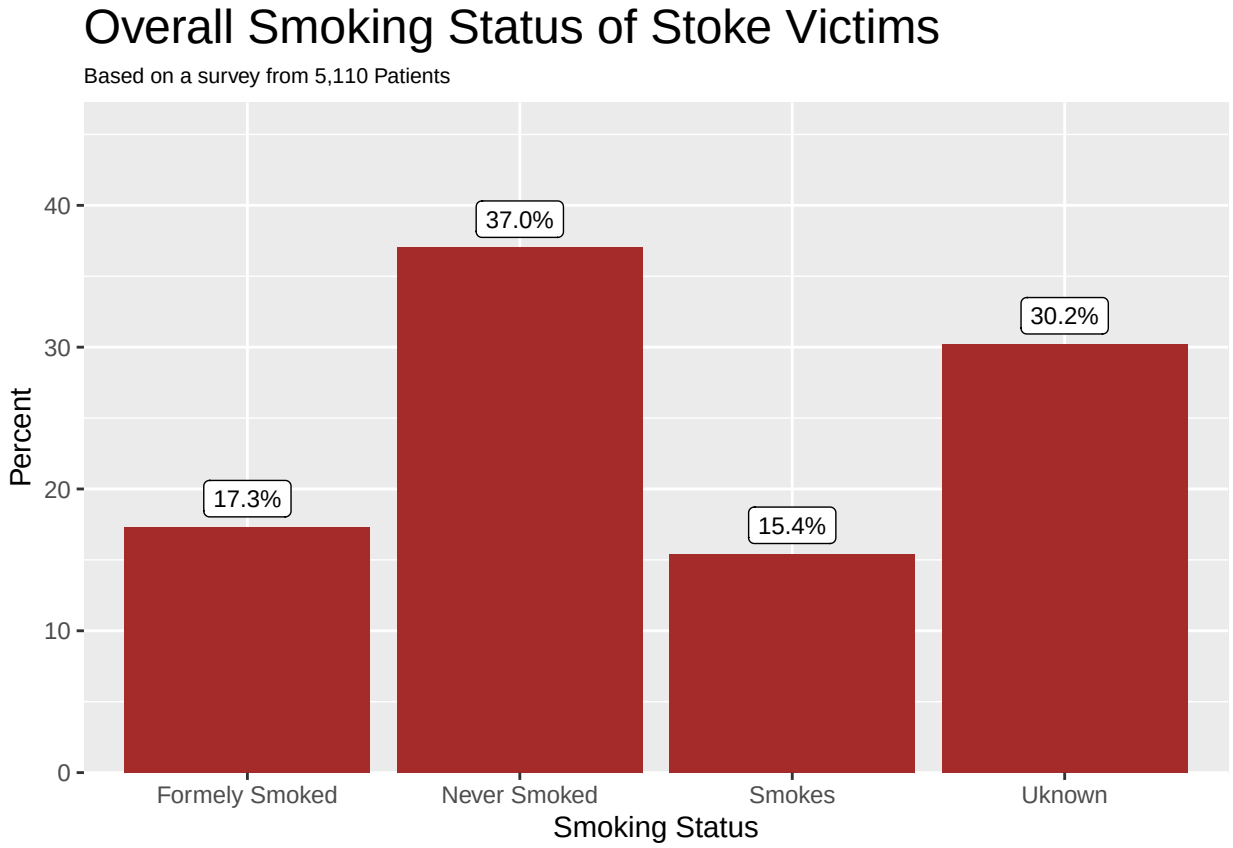
```
attach(health)

smoking<- health %>%
  select(smoking_status)%>%
  count(smoking_status)%>%mutate(prop=100*n/sum(n))

smoking
```

```
## # A tibble: 4 x 3
##   smoking_status      n  prop
## * <chr>           <int> <dbl>
## 1 formerly smoked   885  17.3
## 2 never smoked     1892  37.0
## 3 smokes            789  15.4
## 4 Unknown          1544  30.2
```

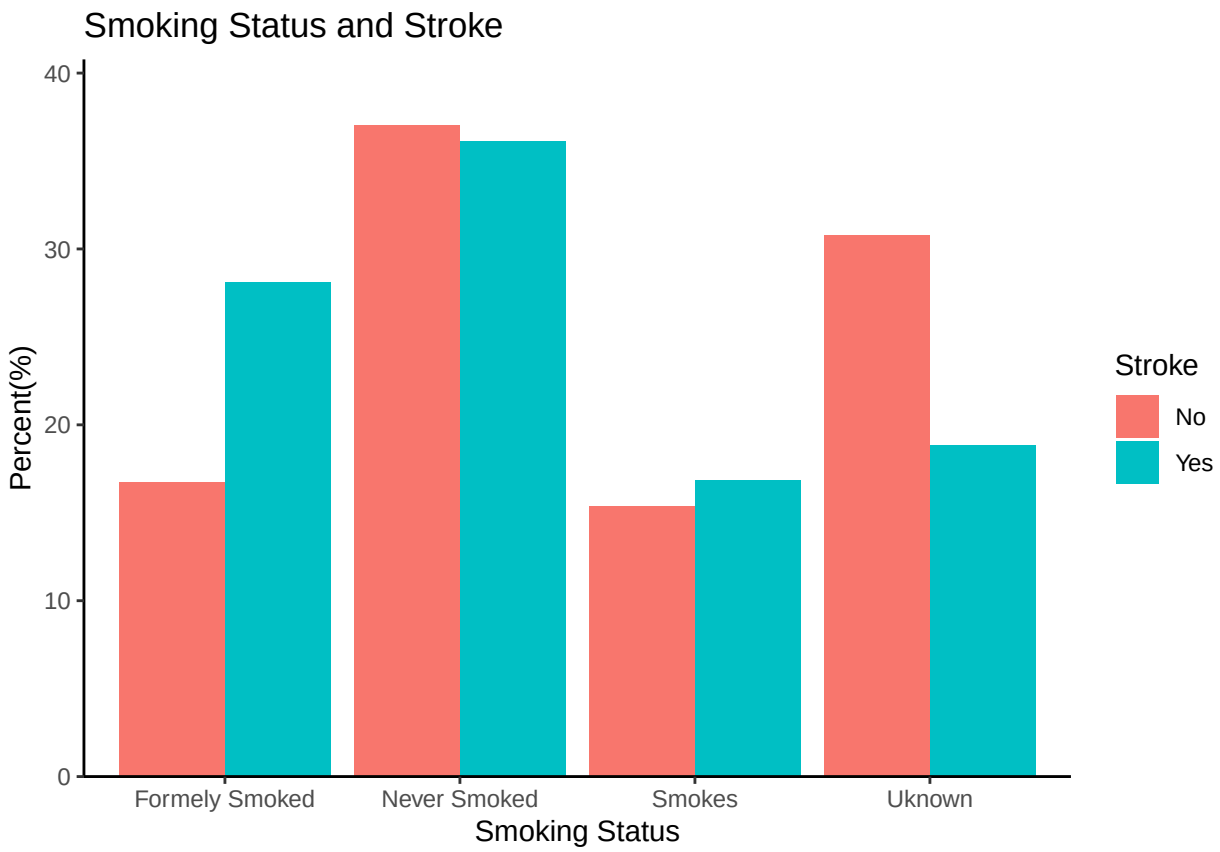
```
ggplot(data=smoking,aes(x=smoking_status, y = prop))+
  geom_bar(stat='identity',fill="brown")+
  scale_y_continuous(expand=expansion(mult=c(0,0.05),
                                      add=c(0,0)),
                     name="Percent", limits=c(0,45))+
  xlab("Smoking Status")+
  ggtitle("Overall Smoking Status of Stoke Victims",
          subtitle="Based on a survey from 5,110 Patients")+
  theme(plot.title=element_text(size = 18),plot.subtitle=element_text(size=8))+
  scale_x_discrete(labels=c("Formely Smoked", "Never Smoked", "Smokes", "Uknown"))+
  geom_label(aes(label=paste0(format(prop, digits=3),"%"), y=prop--2),color="black",size=3.2)
```



Problem 2

```
stroked<-health%>%
mutate(stroke = ifelse(stroke == 1, 'Yes', 'No'))%>%
  group_by(stroke)%>%
 count(smoking_status)%>%
mutate(prop=100*n/sum(n))


ggplot(data=stroked,aes(x=smoking_status, y=prop, fill=stroke))+
  geom_bar(stat="identity", position="dodge")+
  labs(x="Smoking Status", y="Percent(%)", title="Smoking Status and Stroke")+
  theme_classic()+
  scale_y_continuous(expand=expansion(mult=c(0,0.1)))+
  labs(fill = "Stroke")+
  scale_x_discrete(labels=c("Formely Smoked", "Never Smoked", "Smokes", "Uknown"))
```

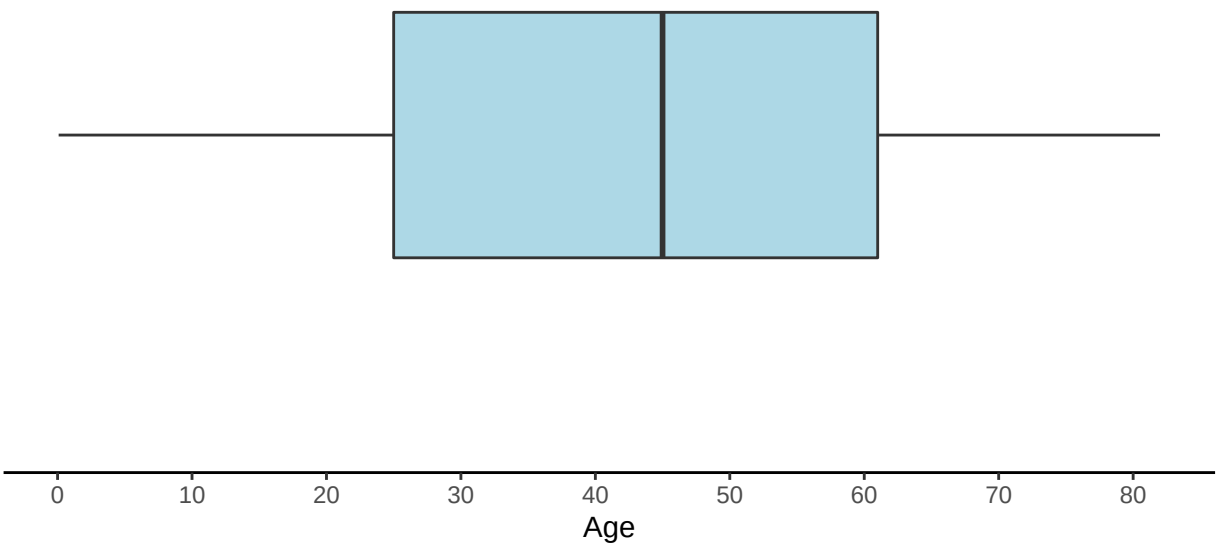


Problem 3

```
ggplot(data = Stroke, aes(x=age)) +
    geom_boxplot(outlier.alpha=0.3,
                 fill="lightblue",
                 outlier.shape = 21,
                 outlier.fill = "lightblue",
                 width=0.2) +
  scale_y_continuous(limits = c(-0.25,0.25)) +
  scale_x_continuous(breaks = seq(0.0,100.0,10.0)) +
  theme_classic() +
  labs(title = "Age Range of the Data",
       subtitle = "Based on Data Predicting Stroke in 5,110 Patients",
       x="Age") +
  theme(axis.line.y = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks.y = element_blank(),
        axis.title.y = element_blank())
```

Age Range of the Data

Based on Data Predicting Stroke in 5,110 Patients



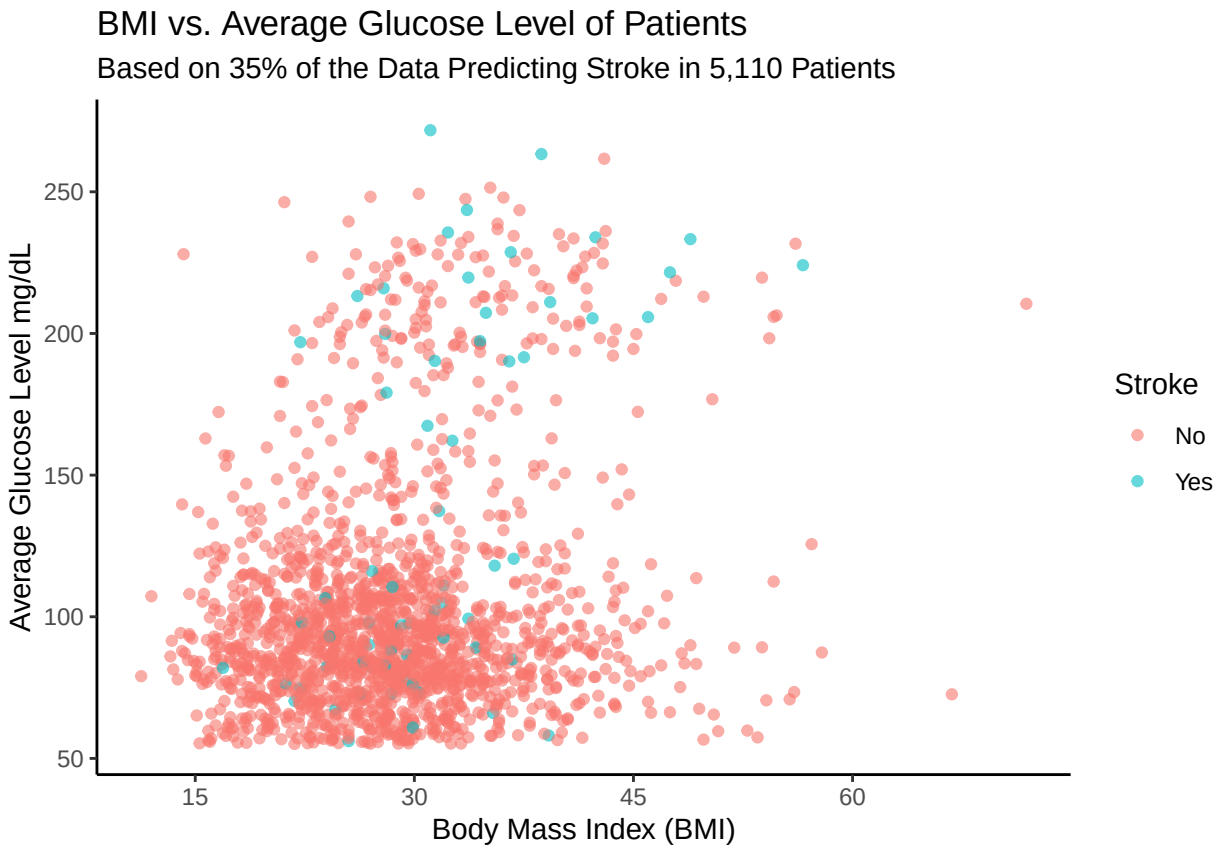

```
#ggsave("Norton_P3.png")
```

Problem 4

```
data <- Stroke %>% replace_with_na(replace = list(bmi = c("N/A"),
                                                  avg_glucose_level = c("N/A")))
```

```
data2 <- data %>% sample_frac(0.35)
#data2
```

```
ggplot(na.omit(data2),
       aes(x=as.numeric(bmi),
           y=as.numeric(avg_glucose_level),
           color=as.factor(stroke))) +
    geom_point(alpha = 0.6) +
    scale_x_continuous(breaks = seq(0.0,100.0,15.0)) +
    labs(title = "BMI vs. Average Glucose Level of Patients",
       subtitle = "Based on 35% of the Data Predicting Stroke in 5,110 Patients",
       y = "Average Glucose Level mg/dL",
       x = "Body Mass Index (BMI)") +
    scale_color_discrete(name="Stroke", labels = c("No","Yes")) +
  theme_classic()
```




```
#ggsave("Norton_P4.png")
```

Problem 5

```
Stroke <- read.csv('../../Data/healthcare-dataset-stroke-data.csv')

min_max_norm <- function(x) {
 (x-min(x, na.rm = TRUE)) / (max(x, na.rm = TRUE)-min(x, na.rm = TRUE))
}

Stroke[,'bmi'] <- as.numeric(Stroke[,'bmi'])
Stroke.sc <-
Stroke %>%
mutate(Stroke_cat = ifelse(stroke == 1, 'Had Stroke', 'No Stroke')) %>%
 select(Stroke_cat,age,hypertension,heart_disease,avg_glucose_level,bmi)


Stroke.sc[,c('age','avg_glucose_level','bmi')] <-
  lapply(Stroke.sc[,c('age','avg_glucose_level','bmi')],min_max_norm)

Stroke.sc <-
Stroke.sc %>% group_by(Stroke_cat) %>%
 summarise(age = mean(age, na.rm = TRUE),
           hypertension = mean(hypertension, na.rm = TRUE),
           heart_disease = mean(heart_disease, na.rm = TRUE),
```

```
            avg_glucose_level = mean(avg_glucose_level, na.rm = TRUE),
            bmi = mean(bmi, na.rm = TRUE))

Stroke.sc <-
Stroke.sc %>% rename(`Glucose level`='avg_glucose_level',
                     BMI = 'bmi',
                     Age = 'age',
                    Hypertension='hypertension',
                     `Heart Disease`='heart_disease')

ggradar(Stroke.sc, base.size = 1,
        plot.extent.x.sf = 1.2,
        plot.title = 'Stroke Associated Factor Comparison',
        legend.position = 'right',
        values.radar = c('0.0','0.5','1.0')) +
 ggsave('RadarPlot.png', width = 200, height = 100, units = 'mm')
```