Introduction to R

Answers to Session 6 exercises

Statistical Consulting Centre

2 March, 2017

1. Plot the mean nerdy score for each gender (with ±1.96×SE bars), as shown in Figure 1.

```
g.m <- with(sports.df, tapply(nerdy.sc, gender, mean, na.rm = TRUE))
g.sd <- with(sports.df, tapply(nerdy.sc, gender, sd, na.rm = TRUE))
g.n <- with(sports.df, tapply(nerdy.sc, gender, function(x)sum(!is.na(x))))
g.stder <- g.sd/sqrt(g.n)

G.df <- data.frame(
  Gender = names(g.m),
  Mean = g.m,
  Upper = g.m + 1.96 * g.stder,
  Lower = g.m - 1.96 * g.stder
  )

ggplot(G.df, aes(x = Gender, y = Mean)) +
  geom_point() +
   geom_errorbar(aes(ymax = Upper, ymin = Lower),
                 width = 0.1)
```

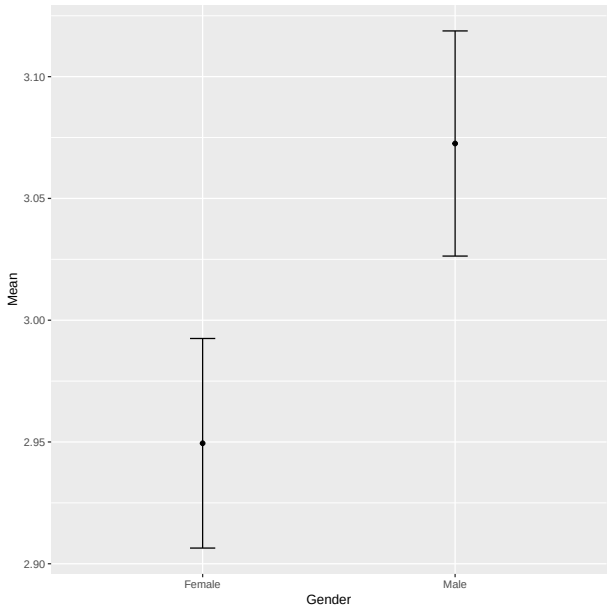

Figure 1: First plot with standard error bars

2. Now reproduce Figure 2. The graph should have:

   - for Males: solid circles (representing the mean) and ±1.96×SE bars (representing the lower and upper 95% confidence limits).
   - for Females: solid circles (representing the mean) and ±1.96×SE bars (representing the lower and upper 95% confidence limits).

```
ga.m <- with(sports.df, tapply(nerdy.sc, list(gender, age.group), mean, na.rm = TRUE))
ga.sd <- with(sports.df, tapply(nerdy.sc, list(gender, age.group), sd, na.rm = TRUE))
ga.n <- with(sports.df, tapply(nerdy.sc, list(gender, age.group), function(x)sum(!is.na(x)
ga.stder <- ga.sd/sqrt(ga.n)
ga.upper <-ga.m + 1.96*ga.stder
ga.lower <-ga.m - 1.96*ga.stder
GA.df = data.frame(
  Age.group = factor(rep(colnames(ga.m), 2), levels = colnames(ga.m)),
  Gender = rep(rownames(ga.m), c(3, 3)),
  Mean = c(ga.m[1, ], ga.m[2, ]),
  Upper = c(ga.upper[1, ], ga.upper[2, ]),
  Lower = c(ga.lower[1, ], ga.lower[2, ])
)
dodge <- position_dodge(width=0.2)
ggplot(GA.df, aes(x = Age.group, y = Mean,
                  color = Gender)) +
  xlab("Age Group")+
  ylab("Mean total Nerdy scores") +
  geom_point(position = dodge) +
  geom_errorbar(aes(ymax = Upper, ymin = Lower),
                width = 0.1, position = dodge)
```

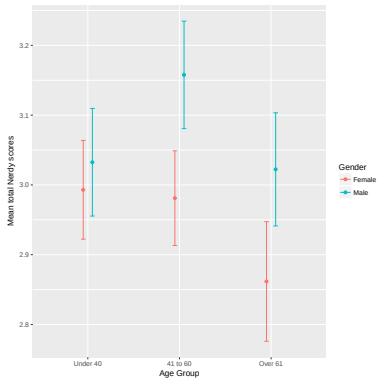

Figure 2: Second plot with standard error bars

3. Now read-in the file NZmap.csv is the Data folder and reproduce Figure 3.

```r
nz1.df <- read.csv("../../Data/NZmap.csv")

nz1.df$Outcome[nz1.df$NAME_1 == "Auckland"] <- 10
nz1.df$Outcome[nz1.df$NAME_1 == "Northland"] <- 11
nz1.df$Outcome[nz1.df$NAME_1 == "Otago"] <- 5
nz1.df$Outcome[nz1.df$NAME_1 == "Southland"] <- 2
nz1.df$Outcome[nz1.df$NAME_1 == "Taranaki"] <- 3
nz1.df$Outcome[nz1.df$NAME_1 == "Waikato"] <- 7
nz1.df$Outcome[nz1.df$NAME_1 == "Wellington"] <- 8
nz1.df$Outcome[nz1.df$NAME_1 == "West Coast"] <- 1
nz1.df$Outcome[nz1.df$NAME_1 == "Bay of Plenty"] <- 9
nz1.df$Outcome[nz1.df$NAME_1 == "Canterbury"] <- 8
nz1.df$Outcome[nz1.df$NAME_1 == "Chatham Islands"] <- 0
nz1.df$Outcome[nz1.df$NAME_1 == "Gisborne"] <- 6
nz1.df$Outcome[nz1.df$NAME_1 == "Hawke's Bay"] <- 6
nz1.df$Outcome[nz1.df$NAME_1 == "Manawatu-Wanganui"] <- 1
nz1.df$Outcome[nz1.df$NAME_1 == "Marlborough"] <- 25
nz1.df$Outcome[nz1.df$NAME_1 == "Nelson"] <- 1

ggplot(data = nz1.df, aes(x = long, y = lat, group = group,
                          fill = Outcome )) +
  geom_polygon(show.legend = TRUE) +
  theme_bw() +
  ylim(4640747, 6255441) +
  xlim( -321280.2, 640000)+
  theme(
    axis.title = element_blank(),
    axis.text = element_blank(),
    axis.ticks = element_blank()
  )
```

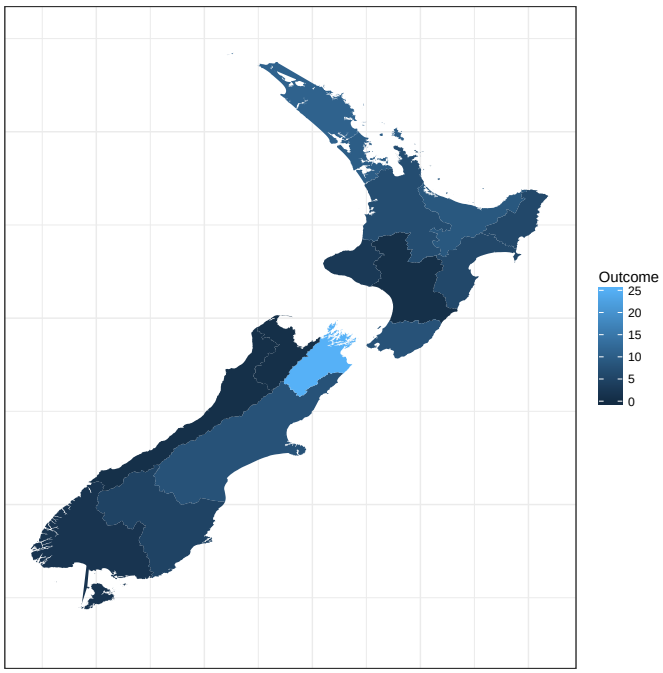

Figure 3: Third plot with mapping