**Name:** Aaditi Rokade **Date:** 10/01/2016

**SCORE:**

**Exercise 1: Concepts, variables, attributes and case summaries**

Use the GSS06 dataset and create case summaries for the variable ‘polviews”:

1. Copy and paste your R output below.

**> unique(GSS2006$polviews)**

**[1] "extremely liberal" "slightly liberal" "conservative"**

**[4] "liberal" "moderate" NA**

**[7] "slghtly conservative" "extrmly conservative”**

**> count(GSS2006, polviews) %>%**

**+ filter(!is.na(polviews)) %>%**

**+ arrange(desc(n)) %>% # Arrange in descending order**

**+ mutate(percent = (n/sum(n)\*100))**

**# A tibble: 7 x 3**

**polviews n percent**

**<chr> <int> <dbl>**

**1 moderate 1683 38.8**

**2 conservative 685 15.8**

**3 slghtly conservative 618 14.3**

**4 liberal 524 12.1**

**5 slightly liberal 517 11.9**

**6 extrmly conservative 167 3.85**

**7 extremely liberal 139 3.21**

What observations can you make?

In 2006, **polviews** (i.e. those who think of self as liberal or conservative) opted from 8 unique options: extremely liberal, sightly liberal, conservative, liberal, moderate, NA, slightly conservative and extremely conservative. As observed from the table with the descending order of number of **polviews,** 38.8% (1683) was the largest sum that opted for ‘**moderate**’ and only 3.21%(139) went with ‘**extremely liberal**’. We can see that conservative is ranked a lot higher than liberal in this table of descending order.

**Exercise 2: Obtaining Graphical Representations of Data in R.**

**Pie Charts:**Use the GSS06 data supplied and create a pie chart for the variable “race”. Copy and paste your R output below;

**> slices\_06 <- c(634, 592, 3284) # inputing the numbers for a slice of the pie chart**

**> lbls\_06 <- c("Black", "Other", "White") # labelng the slices of the pie chart correctly**

**> pct\_06 <- round(slices\_06/sum(slices\_06)\*100) # rounding the numbers and multiplying by 100**

**> lbls\_06 <- paste(lbls\_06, pct\_06) # add percents to labels**

**> lbls\_06 <- paste(lbls\_06,"%",sep="") # add % to labels**

**> pie(slices\_06,labels = lbls\_06, col=rainbow(length(lbls\_06)),**

**+ main="Percent Distribution of Racial Groups - 2006")**

**> pie(slices\_06,labels = lbls\_06, col=rainbow(length(lbls\_06)),**

* **main="Percent Distribution of Racial Groups - 2006”)**

What observations can you make?

As observed from the pie chart for racial distribution (black, white, other). ‘**White**’ contribute to 73% of the overall population in 2006. Whereas ‘**Black**’ contributed to only 14% and others contributed to 13% of the overall population.

**Bar Charts:**Use the GSS06 data supplied and create a bar chart for the variable “conclerg”. What information does this variable capture? Copy and paste your SPSS output below;  **> GSS2006 %>%**

**+ filter(!is.na(conclerg)) %>% # removing all NA's**

**+ ggplot(mapping = aes(x = conclerg, y=..count.., fill = conclerg)) +**

**+ geom\_bar()+**

**+ geom\_text(stat = "count", aes(label=..count..), vjust=1.5) + # adding the raw totals to the bars**

**+ ggtitle("Confidence in Clergy - GSS 2006")+ # main title**

**+ labs(x = "Confidence in Organized Religion") # axis labels**

  What observations can you make?

From the bar chart we observe that, in 2006, conclerg (i.e. confidence in organized religion) was ‘**a great deal**’ for 25.2% of the population, ‘**hardly any**’ for 22.2% populations and ‘**Only some**’ for 52.5% population.

**Histograms:**Use the GSS06 data supplied and create a histogram for the variable “childs”.

What information does this variable capture? Copy and paste your R output below;

**> ggplot(GSS2006, aes(x=childs)) +**

**+ geom\_histogram(binwidth=1.1, colour="black", fill="lightblue") +**

**+ geom\_vline(aes(xintercept=mean(childs, na.rm=T)), # adding line for mean**

**+ color="red", linetype="dashed", size=1) + # Ignore NA**

**+ ggtitle("Number of Children in Household - 2006") + # main title**

**+ labs(x = "Number of Children", y = "Frequency") # axis labels**

What observations can you make?

From the Histogram, we observe that in 2006, the number of children in the house is mostly 0 or 2 for maximum families. Families with either 1 or 3 children are almost similar in number and there are a few families with over 6 children.

**Exercise 2: Obtaining Graphical Representations of Data in R.**

**Pie Charts:**

Use the GSS16 data supplied and create a pie chart for the variable “race”. Copy and paste your R output below;

**> count(GSS2016, race)**

**# A tibble: 3 x 2**

**race n**

**<chr> <int>**

**1 black 490**

**2 other 277**

**3 white 2100**

**> slices\_16 <- c(634, 592, 3284)  > lbls\_16 <- c("Black", "Other", "White")  > pct\_16 <- round(slices\_16/sum(slices\_16)\*100)100**

**> lbls\_16 <- paste(lbls\_16, pct\_16) # add percents to labels**

**> lbls\_16 <- paste(lbls\_16,"%",sep="") # add % to labels**

**> pie(slices\_16,labels = lbls\_16, col=rainbow(length(lbls\_16)),**

* **main="Percent Distribution of Racial Groups - 2016”)     What observations can you make?**As observed from the pie chart for racial distribution (black, white, other). ‘White’ contribute to 73% of the overall population in 2016. Whereas ‘Black’ contributed to only 14% and others contributed to 13% of the overall population. This contribution has particularly remained same over the course of time.

**Bar Charts:**

Use the GSS16 data supplied and create a bar chart for the variable “conclerg”. What information does this variable capture? Copy and paste your R output below;

**> # Bar Chart**

**> count(GSS2016, conclerg) #how did people respond**

**# A tibble: 4 x 2**

**conclerg n**

**<chr> <int>**

**1 a great deal 379**

**2 hardly any 503**

**3 only some 1032**

**4 NA 953**

**> GSS2016 %>% #plots a chart or mapping**

**+ filter(!is.na(conclerg)) %>% # removing all NA's**

**+ ggplot(mapping = aes(x = conclerg, y=..count.., fill = conclerg)) +**

**+ geom\_bar()+**

**+ geom\_text(stat = "count", aes(label=..count..), vjust=1.5) + # adding the raw totals to the bars**

**+ ggtitle("Confidence in Clergy - GSS 2016")+ # main title**

**+ labs(x = "Confidence in Organized Religion") # axis labels**

**> GSS2016 %>% #similar but 2 different instructions for plot**

**+ filter(!is.na(conclerg)) %>% # removing all NA's**

**+ ggplot(mapping = aes(x = conclerg, y=(..count..)/sum(..count..), fill = conclerg)) +**

**+ geom\_bar()+**

**+ geom\_text(stat = "count",**

**+ aes(label = scales::percent((..count..)/sum(..count..))),**

**+ vjust=1.5) + # adding the percentage numbers to the bars**

**+ ggtitle("Percent Distribution of Confidence in Clergy - 2016")+ # main title**

* **labs(x = "Confidence in Organized Religion", y = "Percent") # axis labels       What observations can you make?**

From the bar chart we observe that, in 2016, conclerg (i.e. confidence in organized religion) was ‘**a great deal**’ for 19.8% (decreased as compared to 25.2% in 2006) of the population, ‘**hardly any**’ for 26.3% (increased as compared to 22.2% in 2006) populations and ‘**Only some**’ for 53.9% (slightly increased as compared to 52.5% in 2006) population.

**Histograms:**

Use the GSS16 data supplied and create a histogram for the variable “childs”. What information does this variable capture? Copy and paste your R output below;

**> # Histogram for 'childs'**

**> unique(GSS2016$childs) # unique observations**

**[1] "3" "0" "2" "4"**

**[5] "5" "7" "6" "eight or more"**

**[9] "1" NA**

**> GSS2016$childs <- recode(GSS2016$childs,**

**+ "0" = 0,**

**+ "1" = 1,**

**+ "2" = 2,**

**+ "3" = 3,**

**+ "4" = 4,**

**+ "5" = 5,**

**+ "6" = 6,**

**+ "7" = 7,**

**+ "eight or more" = 8)**

**> unique(GSS2016$childs) # recheck the unique values to**

**[1] 3 0 2 4 5 7 6 8 1 NA**

**> summary(GSS2016$childs) # summary statistics of `childs'**

**Min. 1st Qu. Median Mean 3rd Qu. Max. NA's**

**0.000 0.000 2.000 1.852 3.000 8.000 8**

**> ggplot(GSS2016, aes(x=childs)) +**

**+ geom\_histogram(binwidth=1.1, colour="black", fill="lightblue") +**

**+ geom\_vline(aes(xintercept=mean(childs, na.rm=T)), # adding line for mean**

**+ color="red", linetype="dashed", size=1) + # Ignore NA**

**+ ggtitle("Number of Children in Household - 2016") + # main title**

* **labs(x = "Number of Children", y = "Frequency") # axis labels**

What observations can you make?

From the Histogram above plotted using the GSS data 2016, we observe that the number of children in the house is mostly 0 or 2 for maximum families. Families with either 1 or 3 children are almost similar in number and there are a few families with over 6 children which has remained same as that of GSS data for 2006.