**> #########################################################**

**> ##### R Lab 2: Charts and Graphs -- 2016 GSS Data #####**

**> #########################################################**

**> install.packages("tidyverse")**

**Error in install.packages : Updating loaded packages**

**> install.packages("tidyverse")**

**trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/tidyverse\_1.2.1.tgz'**

**Content type 'application/x-gzip' length 77756 bytes (75 KB)**

**==================================================**

**downloaded 75 KB**

**The downloaded binary packages are in**

**/var/folders/79/tx9tjz8j0hl99904bkw5dy740000gn/T//RtmpJQV4Ht/downloaded\_packages**

**> library(tidyverse)**

**> # Importing GSS 2016 Data**

**> GSS2016 <- read\_csv("/Users/aaditirokade/Desktop/Quant/R\_Lab\_2/GSS\_2016.csv")**

**Parsed with column specification:**

**cols(**

**.default = col\_character(),**

**agekdbrn = col\_integer(),**

**babies = col\_integer(),**

**bigbang = col\_logical(),**

**bigbang1 = col\_logical(),**

**bigbang2 = col\_logical(),**

**boyorgrl = col\_logical(),**

**cohort = col\_integer(),**

**cohrs2 = col\_integer(),**

**condrift = col\_logical(),**

**coninc = col\_double(),**

**conrinc = col\_double(),**

**copres10 = col\_integer(),**

**copres105plus = col\_integer(),**

**cosei10 = col\_double(),**

**cosei10educ = col\_double(),**

**cosei10inc = col\_double(),**

**dateintv = col\_integer(),**

**earnrs = col\_integer(),**

**educ = col\_integer(),**

**electron = col\_logical()**

**# ... with 81 more columns**

**)**

**See spec(...) for full column specifications.**

**Warning: 2 parsing failures.**

**row # A tibble: 2 x 5 col row col expected actual file expected <int> <chr> <chr> <chr> <chr> actual 1 1210 old1 no trailing cha… " or olde… '/Users/aaditirokade/Desktop/Quant… file 2 2760 earnrs an integer eight or … '/Users/aaditirokade/Desktop/Quant…**

**> help("count") #opens in the help tab**

**> help("unique") #Extract Unique Elements**

**> help("mutate")**

**> example("count") #Gives an example**

**count> # tally() is short-hand for summarise()**

**count> mtcars %>% tally()**

**n**

**1 32**

**count> # count() is a short-hand for group\_by() + tally()**

**count> mtcars %>% count(cyl)**

**# A tibble: 3 x 2**

**cyl n**

**<dbl> <int>**

**1 4 11**

**2 6 7**

**3 8 14**

**count> # add\_tally() is short-hand for mutate()**

**count> mtcars %>% add\_tally()**

**# A tibble: 32 x 12**

**mpg cyl disp hp drat wt qsec vs am gear carb n**

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

**1 21 6 160 110 3.9 2.62 16.5 0 1 4 4 32**

**2 21 6 160 110 3.9 2.88 17.0 0 1 4 4 32**

**3 22.8 4 108 93 3.85 2.32 18.6 1 1 4 1 32**

**4 21.4 6 258 110 3.08 3.22 19.4 1 0 3 1 32**

**5 18.7 8 360 175 3.15 3.44 17.0 0 0 3 2 32**

**6 18.1 6 225 105 2.76 3.46 20.2 1 0 3 1 32**

**7 14.3 8 360 245 3.21 3.57 15.8 0 0 3 4 32**

**8 24.4 4 147. 62 3.69 3.19 20 1 0 4 2 32**

**9 22.8 4 141. 95 3.92 3.15 22.9 1 0 4 2 32**

**10 19.2 6 168. 123 3.92 3.44 18.3 1 0 4 4 32**

**# ... with 22 more rows**

**count> # add\_count() is a short-hand for group\_by() + add\_tally()**

**count> mtcars %>% add\_count(cyl)**

**# A tibble: 32 x 12**

**mpg cyl disp hp drat wt qsec vs am gear carb n**

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

**1 21 6 160 110 3.9 2.62 16.5 0 1 4 4 7**

**2 21 6 160 110 3.9 2.88 17.0 0 1 4 4 7**

**3 22.8 4 108 93 3.85 2.32 18.6 1 1 4 1 11**

**4 21.4 6 258 110 3.08 3.22 19.4 1 0 3 1 7**

**5 18.7 8 360 175 3.15 3.44 17.0 0 0 3 2 14**

**6 18.1 6 225 105 2.76 3.46 20.2 1 0 3 1 7**

**7 14.3 8 360 245 3.21 3.57 15.8 0 0 3 4 14**

**8 24.4 4 147. 62 3.69 3.19 20 1 0 4 2 11**

**9 22.8 4 141. 95 3.92 3.15 22.9 1 0 4 2 11**

**10 19.2 6 168. 123 3.92 3.44 18.3 1 0 4 4 7**

**# ... with 22 more rows**

**count> # count and tally are designed so that you can call**

**count> # them repeatedly, each time rolling up a level of detail**

**count> species <- starwars %>% count(species, homeworld, sort = TRUE)**

**count> species**

**# A tibble: 58 x 3**

**species homeworld n**

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

**1 Human Tatooine 8**

**2 Human Naboo 5**

**3 Human NA 5**

**4 Gungan Naboo 3**

**5 Human Alderaan 3**

**6 Droid Tatooine 2**

**7 Droid NA 2**

**8 Human Corellia 2**

**9 Human Coruscant 2**

**10 Kaminoan Kamino 2**

**# ... with 48 more rows**

**count> species %>% count(species, sort = TRUE)**

**# A tibble: 38 x 2**

**species nn**

**<chr> <int>**

**1 Human 16**

**2 Droid 3**

**3 NA 3**

**4 Zabrak 2**

**5 Aleena 1**

**6 Besalisk 1**

**7 Cerean 1**

**8 Chagrian 1**

**9 Clawdite 1**

**10 Dug 1**

**# ... with 28 more rows**

**count> # add\_count() is useful for groupwise filtering**

**count> # e.g.: show only species that have a single member**

**count> starwars %>%**

**count+ add\_count(species) %>%**

**count+ filter(n == 1)**

**# A tibble: 29 x 14**

**name height mass hair\_color skin\_color eye\_color birth\_year gender**

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

**1 Gree… 173 74 NA green black 44 male**

**2 Jabb… 175 1358 NA green-tan… orange 600 herma…**

**3 Yoda 66 17 white green brown 896 male**

**4 Bossk 190 113 none green red 53 male**

**5 Ackb… 180 83 none brown mot… orange 41 male**

**6 Wick… 88 20 brown brown brown 8 male**

**7 Nien… 160 68 none grey black NA male**

**8 Nute… 191 90 none mottled g… red NA male**

**9 Watto 137 NA black blue, grey yellow NA male**

**10 Sebu… 112 40 none grey, red orange NA male**

**# ... with 19 more rows, and 6 more variables: homeworld <chr>, species <chr>,**

**# films <list>, vehicles <list>, starships <list>, n <int>**

**> # We can use unique() to see what are all of the distinct**

**> #observations of the variable we are interested in.**

**> unique(GSS2016$polviews)**

**[1] "moderate" "liberal" "conservative"**

**[4] "slightly liberal" "slghtly conservative" NA**

**[7] "extrmly conservative" "extremely liberal"**

**> count(GSS2016, 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 1032 37.4**

**2 conservative 426 15.5**

**3 slghtly conservative 382 13.9**

**4 liberal 350 12.7**

**5 slightly liberal 310 11.2**

**6 extremely liberal 136 4.93**

**7 extrmly conservative 120 4.35**

**> count(GSS2016, 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 1032 37.4**

**2 conservative 426 15.5**

**3 slghtly conservative 382 13.9**

**4 liberal 350 12.7**

**5 slightly liberal 310 11.2**

**6 extremely liberal 136 4.93**

**7 extrmly conservative 120 4.35**

**> count(GSS2016, 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 1032 37.4**

**2 conservative 426 15.5**

**3 slghtly conservative 382 13.9**

**4 liberal 350 12.7**

**5 slightly liberal 310 11.2**

**6 extremely liberal 136 4.93**

**7 extrmly conservative 120 4.35**

**> count(GSS2016, 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 1032 37.4**

**2 conservative 426 15.5**

**3 slghtly conservative 382 13.9**

**4 liberal 350 12.7**

**5 slightly liberal 310 11.2**

**6 extremely liberal 136 4.93**

**7 extrmly conservative 120 4.35**

**> # We can first get a glimpse of out data**

**> count(GSS2016, race) #gives a count for each race**

**# A tibble: 3 x 2**

**race n**

**<chr> <int>**

**1 black 490**

**2 other 277**

**3 white 2100**

**> # We then use the numebers from the output**

**> # above to create a simple pie chart**

**> slices\_16 <- c(634, 592, 3284) # inputing the numbers for a slice of the pie chart**

**> lbls\_16 <- c("Black", "Other", "White") # labelng the slices of the pie chart correctly**

**> pct\_16 <- round(slices\_16/sum(slices\_16)\*100) # rounding the numbers and multiplying by 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")**

**> # 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**

**> # 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**

**Warning message:**

**Removed 8 rows containing non-finite values (stat\_bin).**