Week 3 Quiz

Ken Wood

7/15/2020

```
# Question #1 - Create a logical vector that identifies the households on greater than 10  acres who s

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv","housing.csv")
housing <- read.csv("housing.csv")
agricultureLogical<-(housing$ACR==3 & housing$AGS==6)
which(agricultureLogical)
```

```
##  [1]  125  238  262  470  555  568  608  643  787  808  824  849  952  955 1033
## [16] 1265 1275 1315 1388 1607 1629 1651 1856 1919 2101 2194 2403 2443 2539 2580
## [31] 2655 2680 2740 2838 2965 3131 3133 3163 3291 3370 3402 3585 3652 3852 3862
## [46] 3912 4023 4045 4107 4113 4117 4185 4198 4310 4343 4354 4448 4453 4461 4718
## [61] 4817 4835 4910 5140 5199 5236 5326 5417 5531 5574 5894 6033 6044 6089 6275
## [76] 6376 6420
```

```
# Question #2 - What are the 30th and 80th quantiles of the resulting data?

library(jpeg)

download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fjeff.jpg","jeff.jpg")
jeff_pic <- readJPEG("jeff.jpg",native=TRUE)
quantile(jeff_pic,probs=c(.3,.8))
```

```
##       30%       80%
## -15259150 -10575416
```

```
# Question #3

download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv","FGDP.csv")
download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv","FEDSTATS_Co

fgdp <- read.csv("FGDP.csv",skip=4,nrows=190)
fgdp <- subset(fgdp, select=-c(X.2,X.5:X.9))
```

```r
colnames(fgdp) <- c("CountryCode","Rank","Country","GDP")

country_stats <- read.csv("FEDSTATS_Country.csv")
```

```r
merged_df <- merge(fgdp,country_stats,by="CountryCode")
nrow(merged_df)
```

```
## [1] 189
```

```r
merged_df <-merged_df[with(merged_df,order(-merged_df$Rank)),]
merged_df[13,3]
```

```
## [1] St. Kitts and Nevis
## 190 Levels: Afghanistan Albania Algeria Angola Antigua and Barbuda ... Zimbabwe
```

```r
# Question #4

high_income_oecd <- merged_df[(merged_df$Income.Group=="High income: OECD"),]
high_income_non_oecd <- merged_df[(merged_df$Income.Group=="High income: nonOECD"),]
x <- mean(as.numeric(high_income_oecd$Rank))
y <- mean(as.numeric(high_income_non_oecd$Rank))
print(c(x,y))
```

```
## [1] 32.96667 91.91304
```

```r
# Question #5 - Cut the GDP ranking into 5 separate quantile groups. Make a table versus Income.Group.

library("Hmisc")
```

```
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
# breaks <- quantile(as.numeric(merged_df$Rank),probs=seq(0,1,0.2),na.rm=TRUE)
# merged_df$quantileGDP <- cut(merged_df$Rank, breaks = breaks)
# result <- merged_df[(merged_df$Income.Group == "Lower middle income" & merged_df$quantileGDP == "(1,3
# result

cutGDP <- cut2(merged_df$Rank, g=5)
table(cutGDP, merged_df$Income.Group)
```

```
##
## cutGDP         High income: nonOECD High income: OECD Low income
##   [  1, 39)  0                    4                18          0
##   [ 39, 77)  0                    5                10          1
```

```
##   [ 77,115)  0                   8                 1          9
##   [115,154)  0                   5                 1         16
##   [154,190]  0                   1                 0         11
##
## cutGDP      Lower middle income Upper middle income
##   [  1, 39)                   5                  11
##   [ 39, 77)                  13                   9
##   [ 77,115)                  12                   8
##   [115,154)                   8                   8
##   [154,190]                  16                   9
```

```
# We see that there are 5 countries that are classified as "lower middle income"
# and are in the top quartile for GDP rank (1,39).
```