quiz_gcd_week1

Derek Hall

5/1/2020

Question 1

How many properties are worth $1,000,000 or more?

```r
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.6.3
```

```r
download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv", destfile = "data/ame

acs_data <- fread("data/american_community_survey.csv")

q1 <- acs_data[VAL == 24, .N]

q1
```

```
## [1] 53
```

Question 3

```r
download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FDATA.gov_NGAP.xlsx", destfile = "

library(xlsx)
```

```
## Warning: package 'xlsx' was built under R version 3.6.3
```

```r
dat <- read.xlsx("data/natgas_aquisition_prog.xlsx", 1, rowIndex = c(18:23), colIndex = c(7:15))

sum(dat$Zip*dat$Ext,na.rm=T)
```

```
## [1] 36534720
```

Question 4

```r
library(XML)
```

```
## Warning: package 'XML' was built under R version 3.6.3
```

```r
library(RCurl)

fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Frestaurants.xml"

# We cannot use htmlTreeParse since it is https therefore use curl and getURL instead
```

```r
# Instead we need to download it using getURL
xmlFile <- getURL(fileURL)

# And then parse the data into doc variable, we need to use htmlParse since this
# is an html file, but it contains embedded XML. useInternal = TRUE gets us all the
# different nodes inside that file
doc <- htmlParse(xmlFile, useInternal = TRUE)

rootNode <- xmlRoot(doc)

zip <- xpathSApply(rootNode, "//zipcode", xmlValue)

sum(zip == 21231)
```

[1] 127
```
```