Week 1 Quiz

Contents

1 Week 1 Quiz

```r
str(download.file)
```

```
## function (url, destfile, method, quiet = FALSE, mode = "w", cacheOK = TRUE,
##     extra = getOption("download.file.extra"), headers = NULL, ...)
```

```r
csvUrl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv"
pdfUrl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FPUMSDataDict06.pdf"
download.file(csvUrl, destfile = "./data/UScommunities.csv", method = "curl")
download.file(pdfUrl, destfile = "./data/UScommunities.pdf", method = "curl")
list.files("./data")
```

```
## [1] "cameras.csv"                        "gas.csv"
## [3] "getdata_data_DATA.gov_NGAP.xlsx" "getdata_data_ss06pid.csv"
## [5] "simple.xml"                         "UScommunities.csv"
## [7] "UScommunities.pdf"
```

```r
uscom <- read.csv("data/UScommunities.csv")
head(uscom)
```

[illegible]																																											
H	186	8	700	4	16	1015675	89	4	1	1	NA	4	2	2	NA	180	0	2	3	3	600	1	NA	1	1300	1	1	1	9	NA	NA	NA	1	1	NA	17	3	840	5	2	105600	NA	NA
H	306	8	700	4	16	1015675	310	1	1	NA	NA	1	7	NA	NA	60	0	2	3	3	NA	1	NA	NA	NA	NA	NA	NA	1	2	2	600	NA	1	3	NA	NA	1	1	3	NA	NA	NA
H	395	8	100	4	16	1015675	106	2	1	1	NA	3	2	2	NA	70	0	2	30	1	200	1	NA	NA	NA	NA	NA	3	1	7	NA	NA	NA	1	2	NA	18	2	50	5	7	9400	NA
H	506	8	700	4	16	1015675	240	4	1	1	NA	4	2	2	NA	40	0	2	80	1	200	1	NA	1	860	1	1	1	6	NA	NA	400	1	1	NA	19	3	500	2	1	66000	NA	NA
H	835	8	800	4	16	1015675	118	4	1	2	1	5	2	2	NA	250	0	2	3	3	700	1	NA	1	1900	1	1	1	7	NA	NA	650	1	1	NA	20	5	2	3	1	93000	NA	NA
H	989	8	700	4	16	1015675	115	4	1	1	NA	3	2	2	NA	130	0	2	3	3	250	1	NA	1	700	1	1	1	6	NA	NA	400	1	1	NA	15	2	1200	5	2	61000	NA	NA

```r
str(uscom)
```

```
## 'data.frame':    6496 obs. of  188 variables:
##  $ RT      : chr  "H" "H" "H" "H" ...
##  $ SERIALNO: int  186 306 395 506 835 989 1861 2120 2278 2428 ...
##  $ DIVISION: int  8 8 8 8 8 8 8 8 8 8 ...
##  $ PUMA    : int  700 700 100 700 800 700 700 200 400 500 ...
##  $ REGION  : int  4 4 4 4 4 4 4 4 4 4 ...
##  $ ST      : int  16 16 16 16 16 16 16 16 16 16 ...
##  $ ADJUST  : int  1015675 1015675 1015675 1015675 1015675 1015675 1015675 1015675 1015675 1015675 ..
##  $ WGTP    : int  89 310 106 240 118 115 0 35 47 51 ...
##  $ NP      : int  4 1 2 4 4 4 1 1 2 2 ...
##  $ TYPE    : int  1 1 1 1 1 1 2 1 1 1 ...
```

```
##  $ ACR     : int  1 NA 1 1 2 1 NA 1 1 1 ...
##  $ AGS     : int  NA NA NA NA 1 NA NA NA NA NA ...
##  $ BDS     : int  4 1 3 4 5 3 NA 2 3 2 ...
##  $ BLD     : int  2 7 2 2 2 2 NA 1 2 1 ...
##  $ BUS     : int  2 NA 2 2 2 2 NA 2 2 2 ...
##  $ CONP    : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ ELEP    : int  180 60 70 40 250 130 NA 40 2 20 ...
##  $ FS      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FULP    : int  2 2 2 2 2 2 NA 480 2 2 ...
##  $ GASP    : int  3 3 30 80 3 3 NA 3 3 140 ...
##  $ HFL     : int  3 3 1 1 3 3 NA 4 3 1 ...
##  $ INSP    : int  600 NA 200 200 700 250 NA NA 770 120 ...
##  $ KIT     : int  1 1 1 1 1 1 NA 1 1 1 ...
##  $ MHP     : int  NA NA NA NA NA NA NA NA NA 220 ...
##  $ MRGI    : int  1 NA NA 1 1 1 NA NA 1 NA ...
##  $ MRGP    : int  1300 NA NA 860 1900 700 NA NA 750 NA ...
##  $ MRGT    : int  1 NA NA 1 1 1 NA NA 1 NA ...
##  $ MRGX    : int  1 NA 3 1 1 1 NA NA 1 3 ...
##  $ PLM     : int  1 1 1 1 1 1 NA 1 1 1 ...
##  $ RMS     : int  9 2 7 6 7 6 NA 4 6 5 ...
##  $ RNTM    : int  NA 2 NA NA NA NA NA NA NA NA ...
##  $ RNTP    : int  NA 600 NA NA NA NA NA NA NA NA ...
##  $ SMP     : int  NA NA NA 400 650 400 NA NA NA NA ...
##  $ TEL     : int  1 1 1 1 1 1 NA 1 1 1 ...
##  $ TEN     : int  1 3 2 1 1 1 NA 4 1 2 ...
##  $ VACS    : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ VAL     : int  17 NA 18 19 20 15 NA NA 13 1 ...
##  $ VEH     : int  3 1 2 3 5 2 NA 1 2 2 ...
##  $ WATP    : int  840 1 50 500 2 1200 NA 650 660 2 ...
##  $ YBL     : int  5 3 5 2 3 5 NA 5 3 5 ...
##  $ FES     : int  2 NA 7 1 1 2 NA NA 2 NA ...
##  $ FINCP   : int  105600 NA 9400 66000 93000 61000 NA NA 209000 NA ...
##  $ FPARC   : int  2 NA 2 1 2 1 NA NA 4 NA ...
##  $ GRNTP   : int  NA 660 NA NA NA NA NA NA NA NA ...
##  $ GRPIP   : int  NA 23 NA NA NA NA NA NA NA NA ...
##  $ HHL     : int  1 1 1 1 1 1 NA 1 1 2 ...
##  $ HHT     : int  1 4 3 1 1 1 NA 6 1 5 ...
##  $ HINCP   : int  105600 34000 9400 66000 93000 61000 NA 10400 209000 35400 ...
##  $ HUGCL   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ HUPAC   : int  2 4 2 1 2 1 NA 4 4 4 ...
##  $ HUPAOC  : int  2 4 2 1 2 1 NA 4 4 4 ...
##  $ HUPARC  : int  2 4 2 1 2 1 NA 4 4 4 ...
##  $ LNGI    : int  1 1 1 1 1 1 NA 1 1 2 ...
##  $ MV      : int  4 3 2 3 1 4 5 5 1 1 ...
##  $ NOC     : int  2 0 1 2 1 2 NA 0 0 0 ...
##  $ NPF     : int  4 NA 2 4 4 4 NA NA 2 NA ...
##  $ NPP     : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ NR      : int  0 0 0 0 0 0 NA 0 0 1 ...
##  $ NRC     : int  2 0 1 2 1 2 NA 0 0 0 ...
##  $ OCPIP   : int  18 NA 23 26 36 26 NA NA 5 7 ...
##  $ PARTNER : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ PSF     : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ R18     : int  1 0 1 1 1 1 NA 0 0 0 ...
##  $ R60     : int  0 0 0 0 0 0 NA 1 1 0 ...
```

```
##  $ R65     : int  0 0 0 0 0 0 NA 1 1 0 ...
##  $ RESMODE : int  1 2 1 2 1 2 NA 2 1 1 ...
##  $ SMOCP   : int  1550 NA 179 1422 2800 1330 NA NA 805 196 ...
##  $ SMX     : int  3 NA NA 1 1 2 NA NA 3 NA ...
##  $ SRNT    : int  0 1 0 0 0 0 NA 1 0 0 ...
##  $ SVAL    : int  1 0 1 1 1 1 NA 0 1 0 ...
##  $ TAXP    : int  24 NA 16 31 25 7 NA NA 22 4 ...
##  $ WIF     : int  3 NA 1 2 3 1 NA NA 1 NA ...
##  $ WKEXREL : int  2 NA 13 2 1 7 NA NA 6 NA ...
##  $ WORKSTAT: int  3 NA 13 1 1 3 NA NA 3 NA ...
##  $ FACRP   : int  0 0 0 0 0 0 NA 0 0 1 ...
##  $ FAGSP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FBDSP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FBLDP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FBUSP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FCONP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FELEP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FFSP    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FFULP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FGASP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FHFLP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FINSP   : int  0 0 0 0 0 1 NA 0 0 0 ...
##  $ FKITP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FMHP    : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FMRGIP  : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FMRGP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FMRGTP  : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FMRGXP  : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FMVYP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FPLMP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FRMSP   : int  0 0 0 0 0 0 NA 0 0 1 ...
##  $ FRNTMP  : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FRNTP   : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FSMP    : int  0 0 0 0 0 0 NA 0 0 0 ...
##  $ FSMXHP  : int  0 0 0 0 0 0 NA 0 0 0 ...
##    [list output truncated]
```

```
# Print count for VAL=24
val_24_count <- val_counts[as.character(24)]
```

```
## Error in eval(expr, envir, enclos): object 'val_counts' not found
```

```
print(val_24_count)
```

```
## Error in eval(expr, envir, enclos): object 'val_24_count' not found
```

```
dat <- read.csv("data/gas.csv",sep = ";")
dat
```

State	AltCoName	Address1	Address2	Address3	City	Zip	CuCurrent	PaCurrent	PuCurrent	Contact	Ext	Fax	email	Status
OK	Tiger Natural Gas	1422 E 71st Suite J	NA		Tulsa	74136	0	1	0	918-491-6998	0	918-491-6659	NA	1

StateAb	CName	Address1	Address2	Address3	City	Zip	CuCurrent	PaCurrent	PoCurrent	Contact	Ext	Fax	email	Status
GA	PS Energy	2987 Clair-mont Rd	Suite 500	NA	Atlanta	30329	1	0	0	404-321-5711	NA		NA	1
OK	Geary Energy	7712 S Yale Ave	Suite 201	NA	Tulsa	74136	1	0	0	918-523-2516	0	918-523-2522	NA	1
CO	Utility Resource Solutions	1700 Lincoln St.	Suite 2530	NA	Denver	80203	0	1	0	303-864-1919	0		NA	1
CO	Select Natural Gas LLC	8122 South-park Lane	Suite 204	NA	Littleton	80120	1	0	0	345-098-8890	456		NA	1

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

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

```r
library(XML)
library(RCurl)
xmlURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Frestaurants.xml"
xmlData <- getURL(xmlURL)
doc <- xmlTreeParse(xmlData, useInternal = TRUE)
```

```
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
```

```
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
## StartTag: invalid element name
```

```
## Error: 1: StartTag: invalid element name
## 2: StartTag: invalid element name
## 3: StartTag: invalid element name
## 4: StartTag: invalid element name
## 5: StartTag: invalid element name
## 6: StartTag: invalid element name
## 7: StartTag: invalid element name
## 8: StartTag: invalid element name
## 9: StartTag: invalid element name
## 10: StartTag: invalid element name
```

```
## 11: StartTag: invalid element name
## 12: StartTag: invalid element name
## 13: StartTag: invalid element name
## 14: StartTag: invalid element name
## 15: StartTag: invalid element name
## 16: StartTag: invalid element name
## 17: StartTag: invalid element name
## 18: StartTag: invalid element name
## 19: StartTag: invalid element name
## 20: StartTag: invalid element name
## 21: StartTag: invalid element name
## 22: StartTag: invalid element name
## 23: StartTag: invalid element name
## 24: StartTag: invalid element name
## 25: StartTag: invalid element name
## 26: StartTag: invalid element name
## 27: StartTag: invalid element name
## 28: StartTag: invalid element name
## 29: StartTag: invalid element name
## 30: StartTag: invalid element name
## 31: StartTag: invalid element name
## 32: StartTag: invalid element name
## 33: StartTag: invalid element name
## 34: StartTag: invalid element name
## 35: StartTag: invalid element name
## 36: StartTag: invalid element name
## 37: StartTag: invalid element name
## 38: StartTag: invalid element name
## 39: StartTag: invalid element name
## 40: StartTag: invalid element name
## 41: StartTag: invalid element name
## 42: StartTag: invalid element name
## 43: StartTag: invalid element name
## 44: StartTag: invalid element name
## 45: StartTag: invalid element name
## 46: StartTag: invalid element name
## 47: StartTag: invalid element name
## 48: StartTag: invalid element name
## 49: StartTag: invalid element name
## 50: StartTag: invalid element name
## 51: StartTag: invalid element name
## 52: StartTag: invalid element name
## 53: StartTag: invalid element name
## 54: StartTag: invalid element name
## 55: StartTag: invalid element name
## 56: StartTag: invalid element name
## 57: StartTag: invalid element name
## 58: StartTag: invalid element name
## 59: StartTag: invalid element name
## 60: StartTag: invalid element name
## 61: StartTag: invalid element name
## 62: StartTag: invalid element name
## 63: StartTag: invalid element name
## 64: StartTag: invalid element name
```

```
## 65: StartTag: invalid element name
## 66: StartTag: invalid element name
## 67: StartTag: invalid element name
## 68: StartTag: invalid element name
## 69: StartTag: invalid element name
## 70: StartTag: invalid element name
## 71: StartTag: invalid element name
```

```r
# load required packages
library(rvest)
library(xml2)

# fetch the XML data
xml_data <- read_xml("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Frestaurants.xml")

# get all the zipcodes
zipcodes <- xml_data %>% xml_find_all("//zipcode") %>% xml_text()

# count the number of restaurants with zipcode 21231
sum(zipcodes == "21231")
```

```
## [1] 127
```

```r
DT  <- fread("data/getdata_data_ss06pid.csv")
```

```
## Error in fread("data/getdata_data_ss06pid.csv"): could not find function "fread"
```

```r
head(DT)
```

```
## Error in eval(expr, envir, enclos): object 'DT' not found
```