STAT243-PS1

Jinhui Xu

September 2017

1 Question 2

a)
download the data from the given url and unzip it.I find that regions of world have "+" in their names.According to this features I can grep and divide data into two parts.

```
wget -O data.zip "http://data.un.org/Handlers/DownloadHandler.ashx?DataFilter=itemCode:
526&DataMartId=FAO&Format=csv&c=2,3,4,5,6,7&s=countryName:asc,elementCode:asc,year:desc"
unzip data.zip
grep "+" data0.csv >regiondata.csv
grep -v "+" data0.csv >countrydata.csv
```

then I deal with country.csv as required.As some countries'name contain "," ,I replace it with blank.In this way,it is convenient to cut it and thus get what I want.Finally use loop to analyse several years.

```
grep "2005" countrydata.csv | grep "Area Harvested" |sed -e 's/, / /g' |sed 's/"//g'|
sort -r -n -t"," -k6 |head -5|cut -d',' -f1
for c in {1965,1975,1985,1995,2005}; do grep "${c}" countrydata.csv | grep "Area Harvested" |
sed -e 's/, / /g' |sed 's/"//g'| sort -r -n -t"," -k6 |head -5|cut -d',' -f1 ;done
```

b)
In myfun(),I first judge whether the numebr of arguments equals to 1 and the judge if the argument is "-h" and else I download the data according to given data codes.As the filename is unknown,I unzip .zip to a new file and rename it and then less it.

```
function myfun(){
  if [ $# -ne "1" ]; then
     echo 'the number of arguments is wrong!'
  elif [ $1 == "-h" ]; then
     echo "download data from http://data.un.org. input product number and get data "
  else
     wget -O data$1.zip "http://data.un.org/Handlers/DownloadHandler.ashx?DataFilter=itemCode:
 $1&DataMartId=FAO&Format=csv&c=2,3,4,5,6,7&s=countryName:asc,elementCode:asc,year:desc"
     unzip -u data$1.zip -d data$1
     mv data$1/* data$1.csv
     rmdir data$1
     rm data$1.zip
    less data$1.csv
  fi
  }
```

2 Question 3

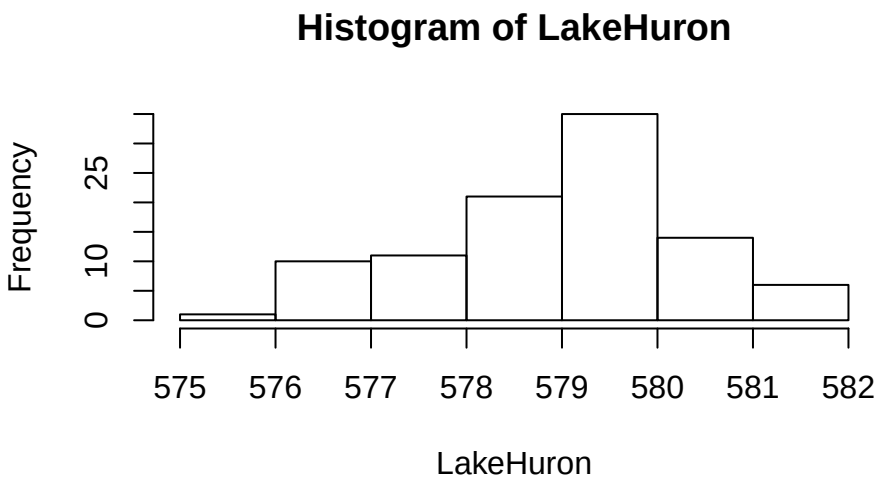
First,I download a .html and get content contains txt via grep; then find location of filename and get it via cut.finally download all txt via loop.

```
wget -O index.html "http://www1.ncdc.noaa.gov/pub/data/ghcn/daily/"
less index.html | grep ".txt" | cut -d "\"" -f8>filename.txt
for file in $(less filename.txt)
do
wget "http://www1.ncdc.noaa.gov/pub/data/ghcn/daily/$file"
echo "successfully download $file"
done
```

3 Question 4

The height of the water level in Lake Huron fluctuates over time. Here I analyze the variation using R. I show a histogram of the lake levels for the period 1875 to 1972

```
hist(LakeHuron)
```




```
lowHi <- c(which.min(LakeHuron), which.max(LakeHuron))
yearExtrema <- attributes(LakeHuron)
```