Problem Set 1

Baoyue Liang

2018/8/31

Problem 3

(a)

Step1: Fetch the temperature data from 2014 to 2018, unzip the zip file

```
cd /Users/lby/Desktop/ps1
for((i=2014; i<=2018; i++)); do
curl -O https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/${i}.csv.gz
gunzip ${i}.csv.gz
echo "year ${i} has $(wc -l) observations"
done;
```

Step2: Count the lines in each csv file, output as observations

```
cd /Users/lby/Desktop/ps1
for ((i=2014; i<=2018; i++)); do
echo "year ${i} has $(wc -l ${i}.csv | cut -d' ' -f2) observations"
done
```

```
## year 2014 has 34599683 observations
## year 2015 has 35233244 observations
## year 2016 has 35384539 observations
## year 2017 has 34748555 observations
## year 2018 has 20100705 observations
```

(b)

Step1: Download the ghcnd-stations.txt file and find "DEATH" in the file to get unique code of Death Valley.
Step2: Find each line that contains the unique code of Death valley, 201X03(which means that the date is in March), and TMAX(which means that temprature is the maximum temprature).

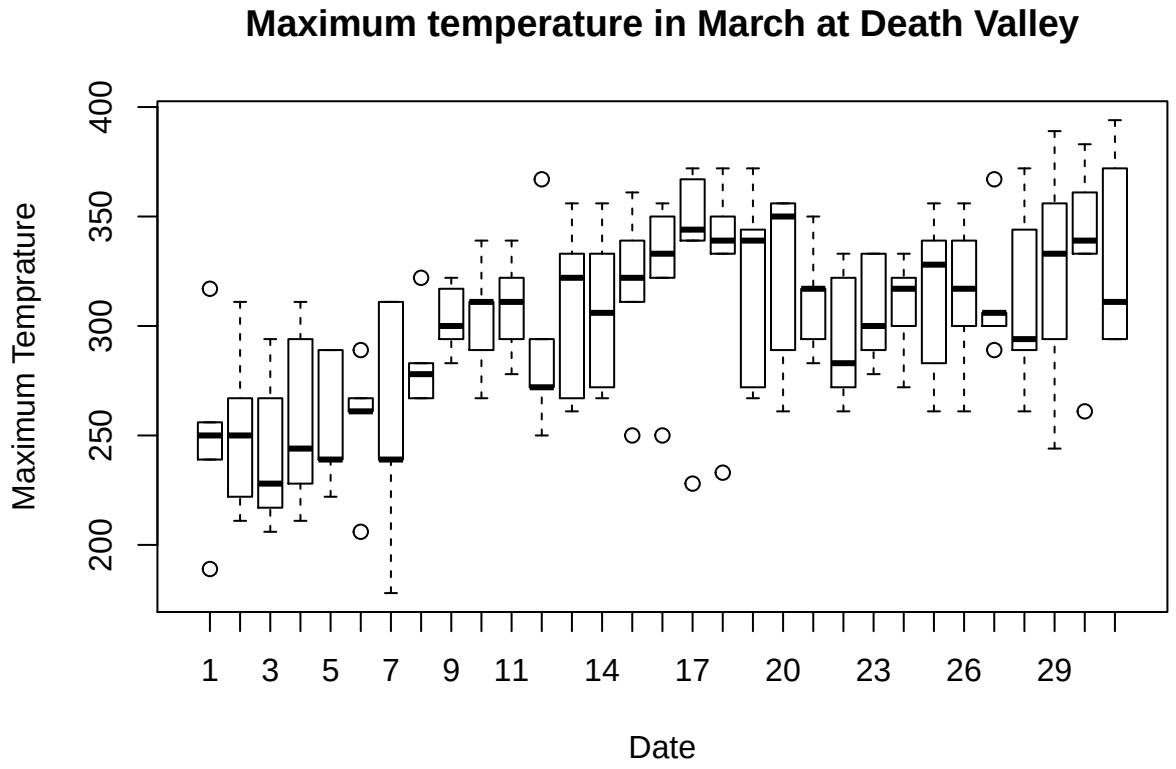
```
cd /Users/lby/Desktop/ps1
curl -o ghcnd-stations.txt https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt
code=$(grep DEATH ghcnd-stations.txt | cut -d' ' -f1)


for((i=2014;i<=2018;i++));do
grep "${code}" ${i}.csv | grep "${i}03" | grep "TMAX" | cut -d',' -f2-4 >> TMAX.txt
done

wc -l TMAX.txt
```

```
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                 Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
```

```
  1 8959k    1  156k    0     0   139k      0  0:01:04  0:00:01  0:01:03  139k
  7 8959k    7  695k    0     0   335k      0  0:00:26  0:00:02  0:00:24  335k
 19 8959k   19 1718k    0     0   552k      0  0:00:16  0:00:03  0:00:13  552k
 35 8959k   35 3195k    0     0   784k      0  0:00:11  0:00:04  0:00:07  784k
 61 8959k   61 5468k    0     0  1069k      0  0:00:08  0:00:05  0:00:03 1091k
 96 8959k   96 8679k    0     0  1429k      0  0:00:06  0:00:06 --:--:-- 1721k
100 8959k  100 8959k    0     0  1468k      0  0:00:06  0:00:06 --:--:-- 2051k
##      155 TMAX.txt
```

(c)

```
setwd("/Users/lby/Desktop/ps1")
TMAX = read.table("TMAX.txt", header = FALSE, sep = ",")
TMAX = TMAX[,-2]
colnames(TMAX) = c("Date","Temprature")
TMAX$Day = TMAX$Date %% 100
boxplot(TMAX$Temprature ~ TMAX$Day, xlab = "Date", ylab = "Maximum Temprature")
title("Maximum temperature in March at Death Valley")
```



(d)

```
cd /Users/lby/Desktop/ps1

function get_weather(){

  location=$1
  wea_var=$2
```

```
  month=$3
  year1=$4
  year2=$5

  if [ "$(echo $1)" == "-h" ] || [ $# != 5 ]; then
  echo "usage: get_weather location weather_type month(double digit) year_to_start year_to_end"
  return
  fi

  if [[ "$year1" -lt "2014" || "$year2" -gt "2018" ]]; then
  echo "No data available. Easy on me."
  return
  fi

  if [ "${year1}" -gt "${year2}" ]; then
  echo -e "Please verify the year you choose. \nCurrently, we do not offer the service to travel back i
  return
  fi

  mon=$(echo $month | wc -c)
  if [[ "$month" -gt "12" || "$mon" -ne "3" ]]; then
  echo -e "Please input a ligit double degit month from 01 to 12. \nThanks for your cooperation."
  return
  fi

  curl -o ghcnd-stations.txt https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt

  code=$(grep -i "${location}" ghcnd-stations.txt | cut -d' ' -f1)

  if [ ! $code ]; then
  echo -e "Unavailable weather station. \nWould you like to check your spelling?"
  return
  fi

  if  [ "$(echo $code | wc -w)" -gt "1" ]; then
  echo -e "Please specify the location. \nJust a kind reminder, the law states that it is illegal to ma
  return
  fi

  for ((i=$year1;i<=$year2;i++)) ;do
  curl -O https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/${i}.csv.gz
  gunzip ${i}.csv.gz
  grep $code ${i}.csv | grep ${i}${month} | grep -i ${wea_var} | cut -d',' -f2-4 >> data.txt
  rm ${i}.csv
  done
  rm ghcnd-stations.txt
  echo "Saved to data.txt, finally, please enjoy the preview"
  head data.txt
}

##Test function

get_weather -h
```

```
echo -e "\n\n"
get_weather a b c d e f g h
echo -e "\n\n"
get_weather "DEATH VALLEY" TMAX 06 2016 2019
echo -e "\n\n"
get_weather "DEATH VALLEY" TMAX 06 2018 2017
echo -e "\n\n"
get_weather "DEATH VALLEY" TMAX 6 2017 2018
echo -e "\n\n"
get_weather "APPLE" TMAX 06 2016 2017
echo -e "\n\n"
get_weather "ASDFG" TMAX 06 2016 2017
echo -e "\n\n"
get_weather "DEATH VALLEY" TMAX 06 2016 2017
```

```
## usage: get_weather location weather_type month(double digit) year_to_start year_to_end
##
##
##
## usage: get_weather location weather_type month(double digit) year_to_start year_to_end
##
##
##
## No data available. Easy on me.
##
##
##
## Please verify the year you choose.
## Currently, we do not offer the service to travel back in time.
##
##
##
## Please input a ligit double degit month from 01 to 12.
## Thanks for your cooperation.
##
##
##
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0 8959k    0 33127    0     0  46345      0  0:03:17 --:--:--  0:03:17 46331
  4 8959k    4  360k    0     0   214k      0  0:00:41  0:00:01  0:00:40  214k
 14 8959k   14 1297k    0     0   477k      0  0:00:18  0:00:02  0:00:16  477k
 30 8959k   30 2719k    0     0   738k      0  0:00:12  0:00:03  0:00:09  738k
 55 8959k   55 5008k    0     0  1061k      0  0:00:08  0:00:04  0:00:04 1061k
 90 8959k   90 8110k    0     0  1428k      0  0:00:06  0:00:05  0:00:01 1627k
100 8959k  100 8959k    0     0  1519k      0  0:00:05  0:00:05 --:--:-- 2040k
## bash: line 35: [: too many arguments
## Please specify the location.
## Just a kind reminder, the law states that it is illegal to marry multiple brides/bridegrooms
## even at different places.
##
##
```

```
##
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                 Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0 8959k    0 14231    0     0  25841      0  0:05:55 --:--:--  0:05:55 25827
  3 8959k    3  352k    0     0   220k      0  0:00:40  0:00:01  0:00:39  220k
 10 8959k   10  946k    0     0   370k      0  0:00:24  0:00:02  0:00:22  370k
 20 8959k   20 1876k    0     0   532k      0  0:00:16  0:00:03  0:00:13  532k
 35 8959k   35 3212k    0     0   702k      0  0:00:12  0:00:04  0:00:08  702k
 56 8959k   56 5032k    0     0   909k      0  0:00:09  0:00:05  0:00:04 1006k
 87 8959k   87 7821k    0     0  1199k      0  0:00:07  0:00:06  0:00:01 1517k
100 8959k  100 8959k    0     0  1301k      0  0:00:06  0:00:06 --:--:-- 1852k
## Unavailable weather station.
## Would you like to check your spelling?
##
##
##
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                 Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0 8959k    0 35263    0     0  24686      0  0:06:11  0:00:01  0:06:10 24676
  3 8959k    3  331k    0     0   139k      0  0:01:04  0:00:02  0:01:02  139k
 11 8959k   11 1034k    0     0   301k      0  0:00:29  0:00:03  0:00:26  301k
 23 8959k   23 2128k    0     0   485k      0  0:00:18  0:00:04  0:00:14  485k
 44 8959k   44 3964k    0     0   735k      0  0:00:12  0:00:05  0:00:07  902k
 76 8959k   76 6878k    0     0  1076k      0  0:00:08  0:00:06  0:00:02 1379k
100 8959k  100 8959k    0     0  1292k      0  0:00:06  0:00:06 --:--:-- 1895k
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                 Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0  192M    0  154k    0     0   130k      0  0:25:08  0:00:01  0:25:07  130k
  0  192M    0  646k    0     0   296k      0  0:11:03  0:00:02  0:11:01  296k
  0  192M    0 1521k    0     0   471k      0  0:06:57  0:00:03  0:06:54  471k
  1  192M    1 2677k    0     0   638k      0  0:05:08  0:00:04  0:05:04  638k
  2  192M    2 4670k    0     0   902k      0  0:03:38  0:00:05  0:03:33  940k
  3  192M    3 7506k    0     0  1211k      0  0:02:42  0:00:06  0:02:36 1467k
  5  192M    5 10.9M    0     0  1565k      0  0:02:05  0:00:07  0:01:58 2119k
  8  192M    8 15.6M    0     0  1960k      0  0:01:40  0:00:08  0:01:32 2930k
 11  192M   11 21.3M    0     0  2383k      0  0:01:22  0:00:09  0:01:13 3852k
 14  192M   14 27.9M    0     0  2808k      0  0:01:10  0:00:10  0:01:00 4782k
 18  192M   18 35.5M    0     0  3259k      0  0:01:00  0:00:11  0:00:49 5806k
 22  192M   22 44.0M    0     0  3704k      0  0:00:53  0:00:12  0:00:41 6774k
 27  192M   27 52.5M    0     0  4080k      0  0:00:48  0:00:13  0:00:35 7551k
 31  192M   31 61.0M    0     0  4406k      0  0:00:44  0:00:14  0:00:30 8120k
 36  192M   36 69.5M    0     0  4689k      0  0:00:41  0:00:15  0:00:26 8517k
 40  192M   40 78.0M    0     0  4937k      0  0:00:39  0:00:16  0:00:23 8689k
 45  192M   45 86.4M    0     0  5156k      0  0:00:38  0:00:17  0:00:21 8692k
 49  192M   49 94.6M    0     0  5334k      0  0:00:36  0:00:18  0:00:18 8637k
 53  192M   53  102M    0     0  5498k      0  0:00:35  0:00:19  0:00:16 8593k
```

```
 57  192M   57  110M    0     0  5615k      0  0:00:35  0:00:20  0:00:15 8428k
 61  192M   61  118M    0     0  5730k      0  0:00:34  0:00:21  0:00:13 8296k
 65  192M   65  126M    0     0  5835k      0  0:00:33  0:00:22  0:00:11 8171k
 70  192M   70  134M    0     0  5959k      0  0:00:33  0:00:23  0:00:10 8234k
 74  192M   74  143M    0     0  6073k      0  0:00:32  0:00:24  0:00:08 8278k
 79  192M   79  151M    0     0  6173k      0  0:00:31  0:00:25  0:00:06 8426k
 83  192M   83  160M    0     0  6271k      0  0:00:31  0:00:26  0:00:05 8557k
 87  192M   87  168M    0     0  6363k      0  0:00:30  0:00:27  0:00:03 8701k
 92  192M   92  177M    0     0  6444k      0  0:00:30  0:00:28  0:00:02 8690k
 96  192M   96  185M    0     0  6522k      0  0:00:30  0:00:29  0:00:01 8695k
100  192M  100  192M    0     0  6576k      0  0:00:29  0:00:29 --:--:-- 8716k
## gunzip: 2016.csv already exists -- skipping
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                 Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0  189M    0 57116    0     0  65912      0  0:50:10 --:--:--  0:50:10 65877
  0  189M    0  493k    0     0   268k      0  0:12:02  0:00:01  0:12:01  268k
  0  189M    0 1508k    0     0   524k      0  0:06:09  0:00:02  0:06:07  524k
  1  189M    1 3118k    0     0   812k      0  0:03:58  0:00:03  0:03:55  812k
  2  189M    2 5563k    0     0  1139k      0  0:02:50  0:00:04  0:02:46 1139k
  4  189M    4 8860k    0     0  1516k      0  0:02:07  0:00:05  0:02:02 1769k
  7  189M    7 13.3M    0     0  1994k      0  0:01:37  0:00:06  0:01:31 2629k
 10  189M   10 19.8M    0     0  2592k      0  0:01:14  0:00:07  0:01:07 3791k
 14  189M   14 27.4M    0     0  3177k      0  0:01:00  0:00:08  0:00:52 4993k
 18  189M   18 35.6M    0     0  3709k      0  0:00:52  0:00:09  0:00:43 6241k
 23  189M   23 44.1M    0     0  4170k      0  0:00:46  0:00:10  0:00:36 7271k
 27  189M   27 52.6M    0     0  4552k      0  0:00:42  0:00:11  0:00:31 8053k
 32  189M   32 61.1M    0     0  4876k      0  0:00:39  0:00:12  0:00:27 8456k
 36  189M   36 69.6M    0     0  5152k      0  0:00:37  0:00:13  0:00:24 8643k
 41  189M   41 78.1M    0     0  5391k      0  0:00:35  0:00:14  0:00:21 8700k
 45  189M   45 86.6M    0     0  5600k      0  0:00:34  0:00:15  0:00:19 8700k
 50  189M   50 95.1M    0     0  5784k      0  0:00:33  0:00:16  0:00:17 8700k
 54  189M   54  103M    0     0  5936k      0  0:00:32  0:00:17  0:00:15 8654k
 58  189M   58  111M    0     0  6065k      0  0:00:31  0:00:18  0:00:13 8593k
 63  189M   63  120M    0     0  6198k      0  0:00:31  0:00:19  0:00:12 8593k
 67  189M   67  128M    0     0  6316k      0  0:00:30  0:00:20  0:00:10 8584k
 72  189M   72  137M    0     0  6425k      0  0:00:30  0:00:21  0:00:09 8584k
 76  189M   76  145M    0     0  6525k      0  0:00:29  0:00:22  0:00:07 8630k
 81  189M   81  153M    0     0  6611k      0  0:00:29  0:00:23  0:00:06 8670k
 85  189M   85  162M    0     0  6696k      0  0:00:28  0:00:24  0:00:04 8671k
 90  189M   90  170M    0     0  6773k      0  0:00:28  0:00:25  0:00:03 8681k
 94  189M   94  179M    0     0  6835k      0  0:00:28  0:00:26  0:00:02 8626k
 99  189M   99  187M    0     0  6897k      0  0:00:28  0:00:27  0:00:01 8598k
100  189M  100  189M    0     0  6912k      0  0:00:28  0:00:28 --:--:-- 8622k
## gunzip: 2017.csv already exists -- skipping
## Saved to data.txt, finally, please enjoy the preview
## 20160601,TMAX,467
## 20160602,TMAX,489
## 20160603,TMAX,478
## 20160604,TMAX,489
## 20160605,TMAX,483
## 20160606,TMAX,461
## 20160607,TMAX,478
```

```
## 20160608,TMAX,489
## 20160609,TMAX,478
## 20160610,TMAX,444
```

Problem 4

First, we should get the url and select all the file name of the txt files. Then, we append the txt file name after the address to download the file and echo the name of the file.

```
cd /Users/lby/Desktop/ps1
curl 'https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/'| grep -o -E '"\S*.txt"' > name.txt

for URL in `cat name.txt`; do
curl -O https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/${URL};
echo "${URL}";
done;
```

```
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100  6068  100  6068    0     0  11673      0 --:--:-- --:--:-- --:--:-- 11691
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0   249    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100   249  100   249    0     0    565      0 --:--:-- --:--:-- --:--:--   564
## "ghcnd-countries.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100   249  100   249    0     0    558      0 --:--:-- --:--:-- --:--:--   559
## "ghcnd-inventory.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100   246  100   246    0     0    537      0 --:--:-- --:--:-- --:--:--   538
## "ghcnd-states.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100   248  100   248    0     0    549      0 --:--:-- --:--:-- --:--:--   549
## "ghcnd-stations.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100   247  100   247    0     0    551      0 --:--:-- --:--:-- --:--:--   551
```

```
## "ghcnd-version.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100   245  100   245    0     0    549      0 --:--:-- --:--:-- --:--:--   549
## "mingle-list.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100   240  100   240    0     0    495      0 --:--:-- --:--:-- --:--:--   495
## "readme.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100   240  100   240    0     0    541      0 --:--:-- --:--:-- --:--:--   541
## "status.txt"
```

Problem 5

(b)

```
library(reticulate)
x = c(1,2,3)
r_to_py(x, convert = FALSE)
```

```
## [1.0, 2.0, 3.0]
```