UFO_R_Markdown

Carlos

2022-09-06

The first step when running a code in R is to install the necessary packages. In this case we will be using the tidyverse package. To install the latest version of this package we run the chunk below:

```
install.packages("tidyverse",repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/simpl/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\simpl\AppData\Local\Temp\Rtmp88cE5A\downloaded_packages
```

We then load the tidyverse package

```
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Data Import

We will then need to import the data we have calculated in SQL we can do this with the read_csv() function

```
observations_per_country <-read.csv("https://raw.githubusercontent.com/Carly1269/PortfolioProjects/main/

observations_per_hour_of_the_day <-read.csv("https://raw.githubusercontent.com/Carly1269/PortfolioProje

observations_per_state <-read.csv("https://raw.githubusercontent.com/Carly1269/PortfolioProjects/main/U

observations_per_year <-read.csv("https://raw.githubusercontent.com/Carly1269/PortfolioProjects/main/UF
```

We then print the data to check on the data types

```
print(observations_per_country)
```

```
##   county_sighting_count country_clean
## 1                 70927           us
## 2                  5969          N/A
## 3                  3436           ca
```

```
print(observations_per_hour_of_the_day)
```

```
##    hour_observed number_of_observations
## 1             21                  11445
## 2             22                  10837
## 3             20                   8617
## 4             23                   7953
## 5             19                   6147
## 6              0                   4108
## 7             18                   4002
## 8              1                   3210
## 9             17                   2592
## 10             2                   2357
## 11             3                   2004
## 12            16                   1620
## 13             5                   1591
## 14             4                   1529
## 15            15                   1433
## 16            12                   1368
## 17            14                   1322
## 18            13                   1303
## 19             6                   1224
## 20            10                   1166
## 21            11                   1144
## 22             9                    958
## 23             7                    905
## 24             8                    803
## 25            NA                    694
```

```
print(observations_per_state)
```

```
##    sate_sighting_count stateid
## 1                 9655      ca
## 2                 4266      wa
## 3                 4200      fl
## 4                 3677      tx
## 5                 3219      ny
## 6                 2689      az
## 7                 2645      il
## 8                 2582      pa
## 9                 2424      oh
## 10                2071      mi
## 11                1868      nc
## 12                1845      or
```

```
## 13              1584      on
## 14              1576      mo
## 15              1512      nj
## 16              1505      co
## 17              1393      va
## 18              1386      in
## 19              1358      ma
## 20              1347      ga
## 21              1333      wi
## 22              1192      tn
## 23              1081      mn
## 24              1076      sc
## 25               968      ct
## 26               914      ky
## 27               911      md
## 28               905      nv
## 29               815      nm
## 30               787      bc
## 31               766      ok
## 32               743      ut
## 33               707      ia
## 34               690      al
## 35               666      ar
## 36               653      ks
## 37               633      me
## 38               597      la
## 39               554      id
## 40               535      nh
## 41               510      mt
## 42               485      wv
## 43               414      ms
## 44               405      ne
## 45               354      ak
## 46               353      hi
## 47               333      ab
## 48               307      vt
## 49               289      ri
## 50               205      wy
## 51               196      sd
## 52               183      de
## 53               178      qc
## 54               155      mb
## 55               141      ns
## 56               138      nd
## 57               116      nb
## 58                98      dc
## 59                97      sk
## 60                90      pq
## 61                33      pr
## 62                28      sa
## 63                25      nf
## 64                17      pe
## 65                17      nt
## 66                11      yt
```

```
## 67                   7      yk
## 68                   1
```

```
print(observations_per_year)
```

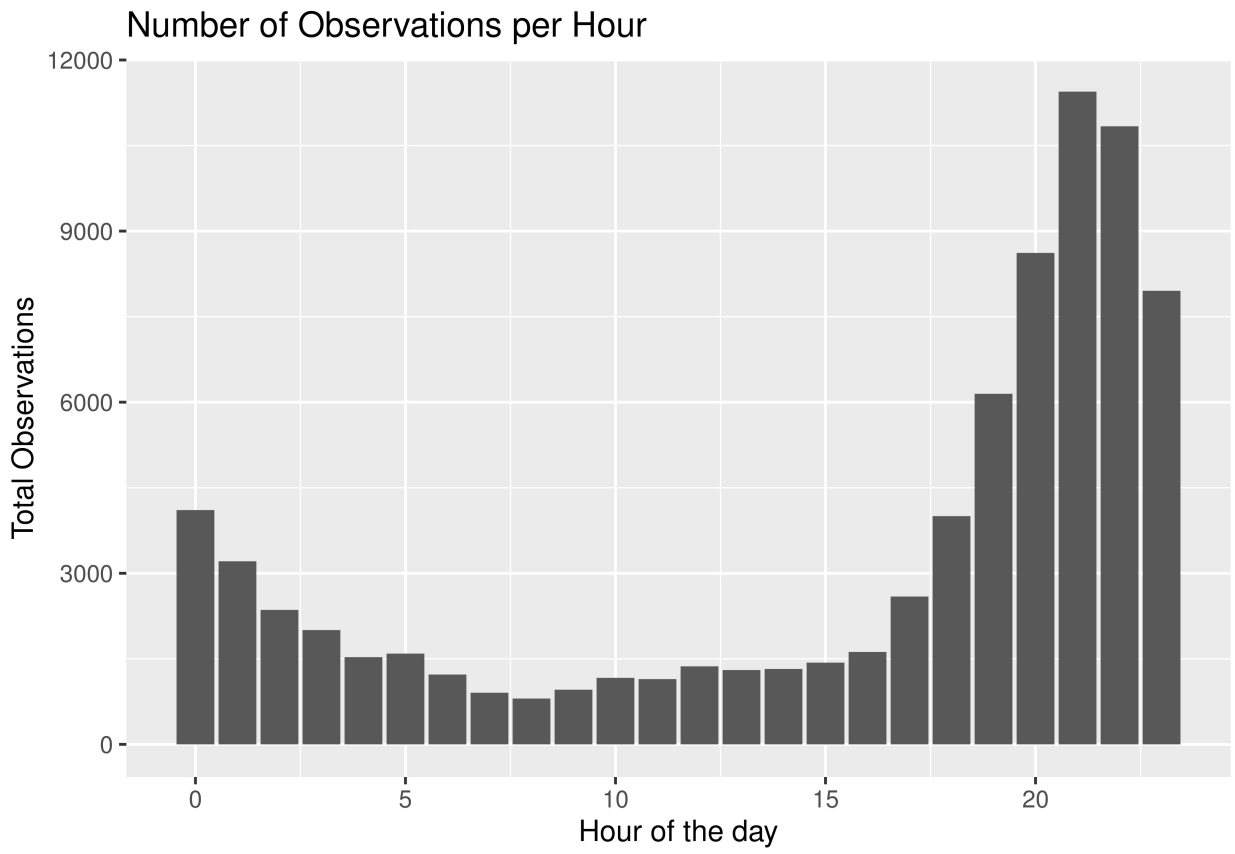
```
##    year_observed number_of_observations
## 1           2012                   7308
## 2           2013                   6992
## 3           2011                   5077
## 4           2008                   4777
## 5           2009                   4508
## 6           2010                   4258
## 7           2007                   4243
## 8           2004                   4220
## 9           2005                   4039
## 10          2003                   3935
## 11          2006                   3682
## 12          2002                   3211
## 13          2001                   3109
## 14          1999                   2788
## 15          2000                   2752
## 16          2014                   2252
## 17          1998                   1739
## 18          1997                   1242
## 19          1995                   1066
## 20          1996                    826
## 21            -1                    694
## 22          1994                    417
## 23          1978                    330
## 24          1975                    311
## 25          1993                    300
## 26          1976                    273
## 27          1974                    269
## 28          1977                    264
## 29          1990                    257
## 30          1992                    246
## 31          1979                    244
## 32          1989                    243
## 33          1988                    239
## 34          1980                    238
## 35          1991                    234
## 36          1987                    223
## 37          1973                    222
## 38          1985                    215
## 39          1968                    214
## 40          1986                    188
## 41          1966                    185
## 42          1984                    185
## 43          1967                    184
## 44          1965                    184
## 45          1982                    180
## 46          1981                    163
## 47          1969                    155
## 48          1972                    155
```

```
## 49          1983                    148
## 50          1970                    144
## 51          1971                    128
## 52          1964                     88
## 53          1963                     82
## 54          1962                     71
## 55          1957                     70
## 56          1960                     63
## 57          1954                     51
## 58          1959                     49
## 59          1961                     48
## 60          1952                     45
## 61          1958                     44
## 62          1956                     42
## 63          1947                     35
## 64          1955                     31
## 65          1953                     31
## 66          1950                     27
## 67          1951                     20
## 68          1949                     16
## 69          1946                      9
## 70          1945                      9
## 71          1944                      9
## 72          1943                      9
## 73          1948                      7
## 74          1939                      3
## 75          1931                      2
## 76          1936                      2
## 77          1942                      2
## 78          1937                      1
## 79          1906                      1
## 80          1934                      1
## 81          1925                      1
## 82          1916                      1
## 83          1933                      1
## 84          1920                      1
## 85          1930                      1
## 86          1929                      1
## 87          1941                      1
## 88          1910                      1
```

We can see that the data types are correct. We will then proceed with our analysis ## Data Graphing We will first graph the total number observations per hour of the day

```
ggplot(data =observations_per_hour_of_the_day, mapping = aes(x=hour_observed, y=number_of_observations)
```

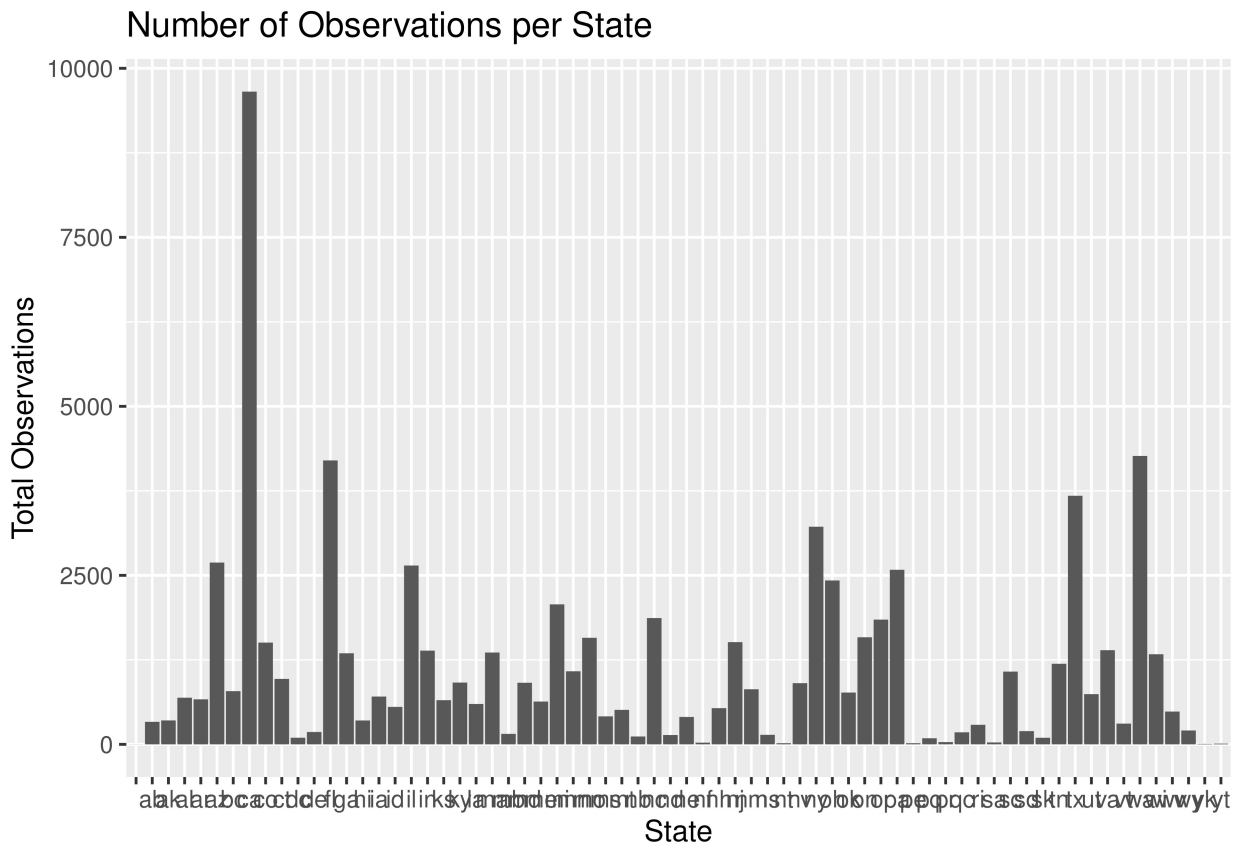
```
## Warning: Removed 1 rows containing missing values (position_stack).
```



As we can see most observations happen during the night time.

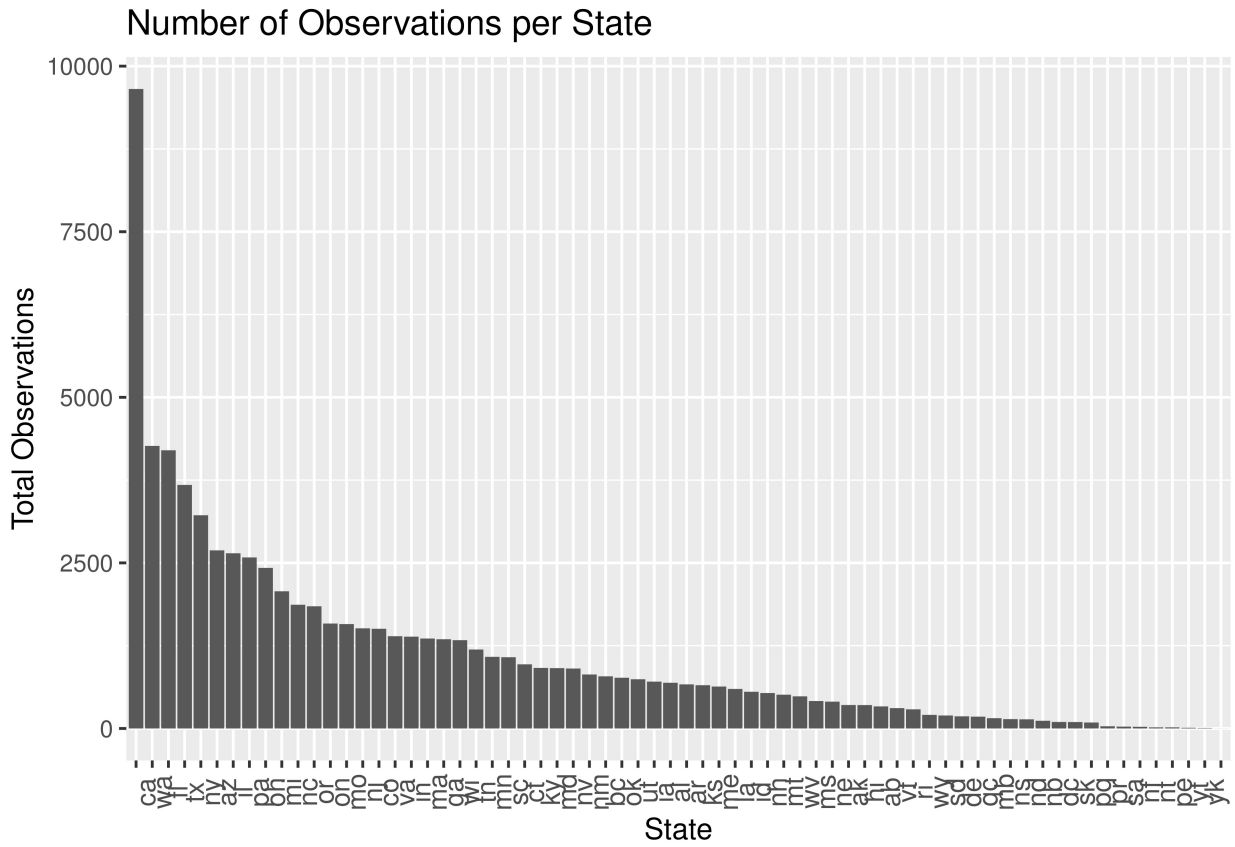
We will then graph the data of the total number of observations per state:

```
ggplot(data =observations_per_state, mapping = aes(x=stateid, y=sate_sighting_count))+ geom_bar(stat =
```



We can see that the data varies a lot. To visualize it better we will order the states from the most number of observations to the least. We will also tilt the state ids for better visualization

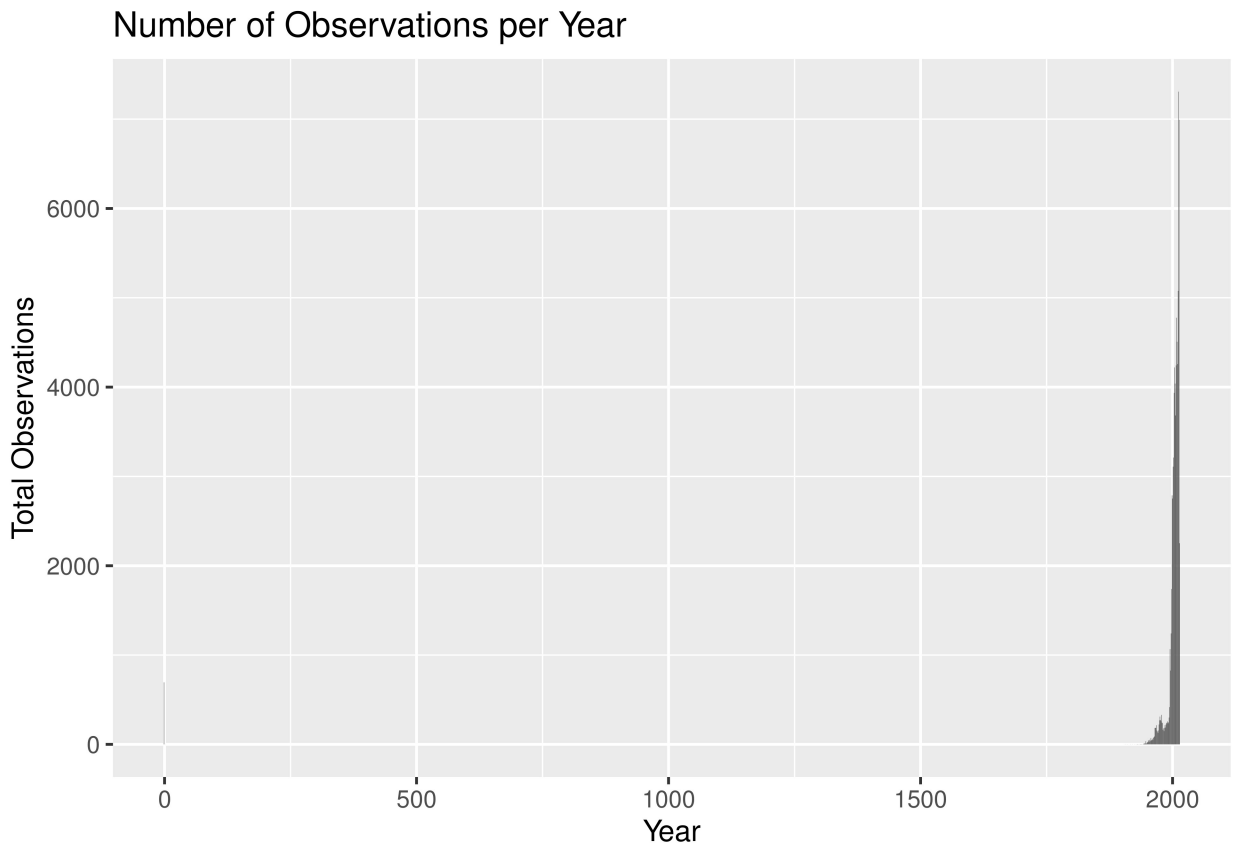
```
ggplot(data =observations_per_state, mapping = aes(x=reorder(stateid,-sate_sighting_count), y=sate_sigh
```



We can see that the state of California has the most number of observations. Followed by the state of Washington

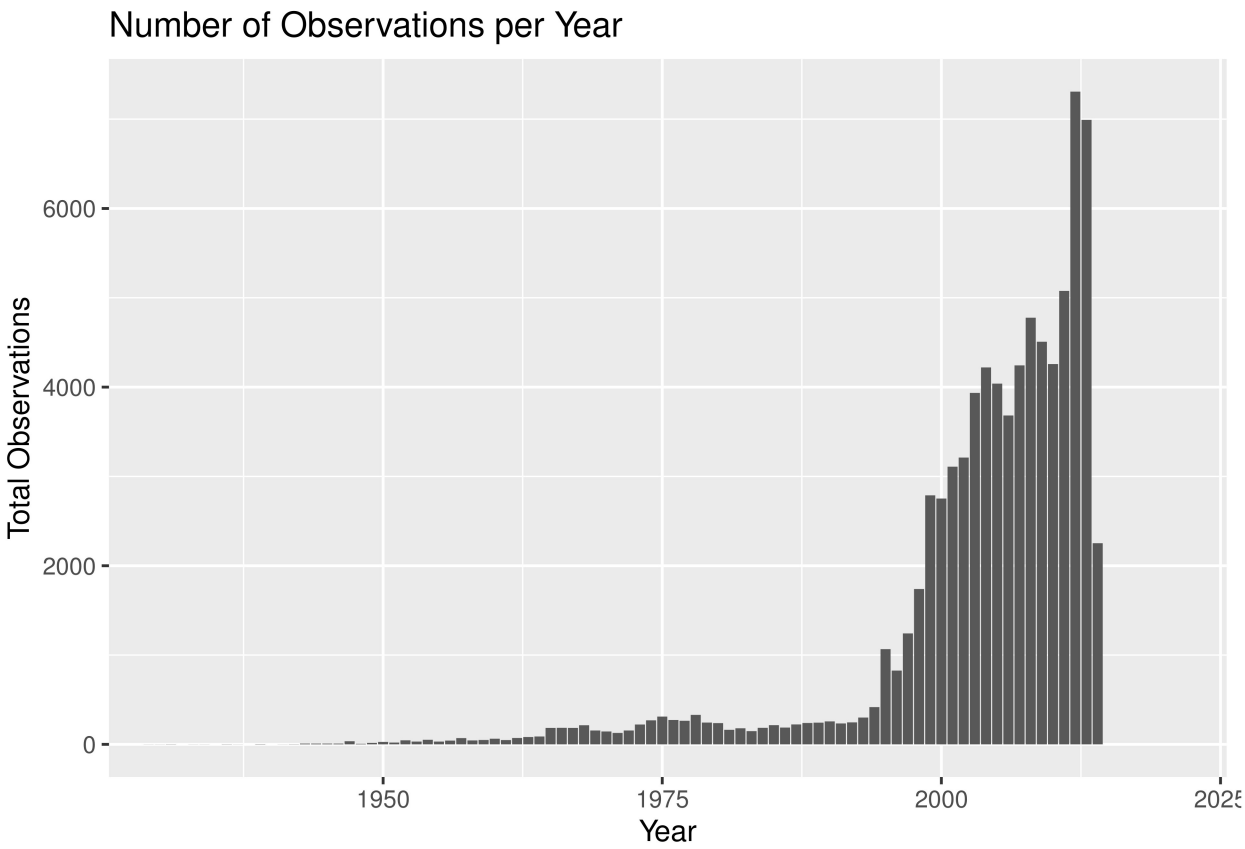
We will then graph the total number of observations per year

```
ggplot(data =observations_per_year, mapping = aes(x=year_observed, y=number_of_observations))+ geom_bar
```



The graph is really hard to visualize since we have observations from thousands of years in the past. We will zoom into our graph to better see our data

```
ggplot(data =observations_per_year, mapping = aes(x=year_observed, y=number_of_observations))+ geom_bar
```


Number of Observations per Year
Total Observations
6000
4000
2000
0
1950
1975
2000
2025
Year