COMPSCIX 415.2 Homework 1

Santosh Kanutala

6/7/2018

My Github repository for my assignments can be found at this URL : (https://github.com/santumagic/compscix-415-2assignments.git)

```r
library(mdsr)
library(tidyverse)

data("WorldCities")

# QUESTION 1 :

glimpse(WorldCities)
```

```
## Observations: 23,018
## Variables: 10
## $ code         <int> 3040051, 3041563, 290594, 291074, 291696, 292223...
## $ name         <chr> "les Escaldes", "Andorra la Vella", "Umm al Qayw...
## $ latitude     <dbl> 42.50729, 42.50779, 25.56473, 25.78953, 25.33132...
## $ longitude    <dbl> 1.53414, 1.52109, 55.55517, 55.94320, 56.34199, ...
## $ country      <chr> "AD", "AD", "AE", "AE", "AE", "AE", "AE", ...
## $ countryRegion <chr> "8", "7", "7", "5", "6", "3", "4", "6", "1", "4"...
## $ population   <dbl> 15853, 20430, 44411, 115949, 33575, 1137347, 263...
## $ regionCode   <int> 1033, 1037, 2, 2, 20, 11, 4, 6, 16, 15, 275, 4, ...
## $ region       <chr> "Europe/Andorra", "Europe/Andorra", "Asia/Dubai"...
## $ date         <chr> "10/15/08", "5/30/10", "11/3/12", "11/30/12", "1...
## In the above WorldCities dataset there are 23,018 observations and 10 Variables.
## Below are some variable names.
```

code, name, lattitude, longitude, country, countryRegion, population, regionCode, region, date

```r
## Extract parts of the data

WorldCities <- head(WorldCities, 200) # Top 200 rows
country_col <- WorldCities$country # Pulling the country column by assigning it to country_col
unique(country_col) # Displaying the unique values from the field country_col
```

```
## [1] "AD" "AE" "AF" "AG" "AI" "AL" "AM" "AO" "AR"
```

```r
# QUESTION 2 :

region_col <- WorldCities$region # Pulling the region column by assigning it to region_col
unique(region_col) # Displaying the unique values from the field region
```

```
##  [1] "Europe/Andorra"              "Asia/Dubai"
##  [3] "Asia/Kabul"                  "America/Antigua"
##  [5] "America/Anguilla"            "Europe/Tirane"
##  [7] "Asia/Yerevan"                "Africa/Luanda"
##  [9] "America/Argentina/Buenos_Aires" "America/Argentina/Cordoba"
## [11] "America/Argentina/Salta"     "America/Argentina/Tucuman"
## [13] "America/Argentina/San_Juan"
```

```
# QUESTION 3 :

country_col <- WorldCities %>% select(country) %>% unique()
country_col %>% unique()
```
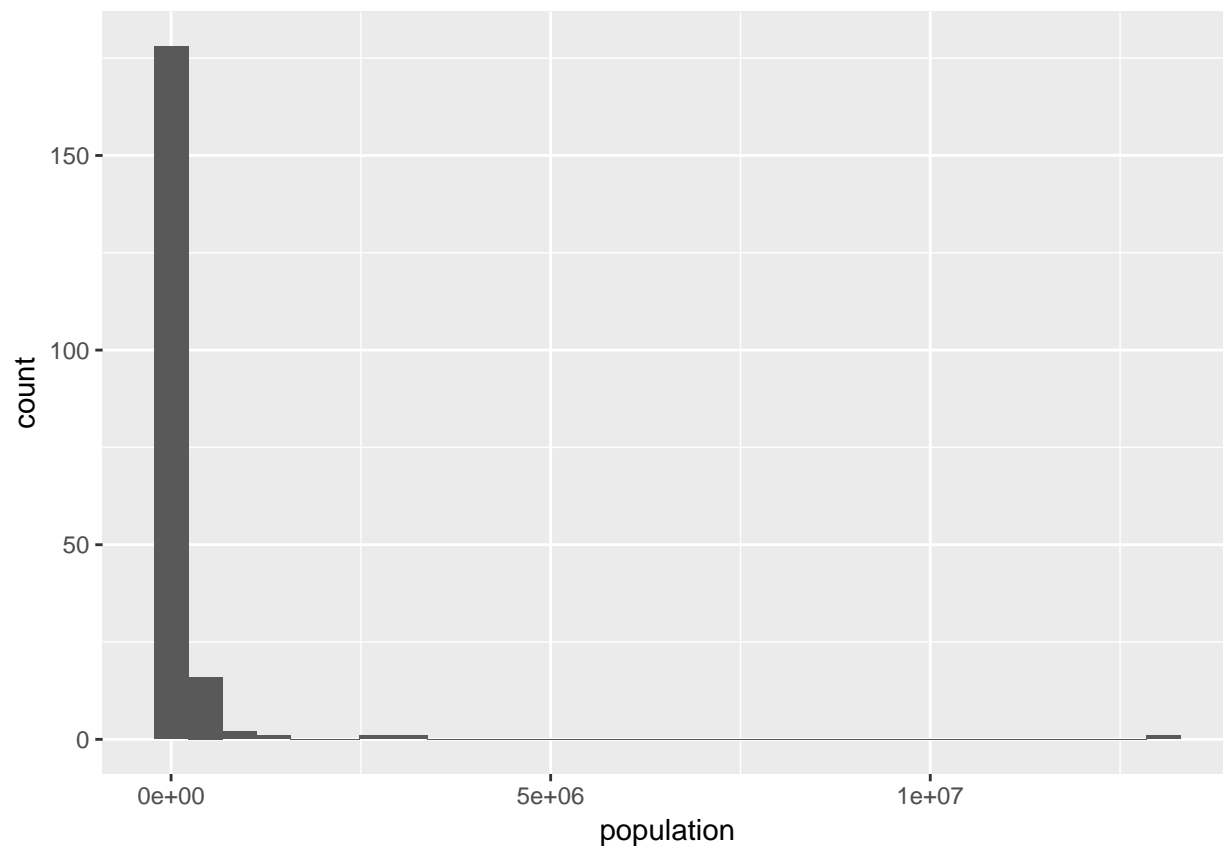
```
##      country
## 1       AD
## 3       AE
## 15      AF
## 65      AG
## 66      AI
## 67      AL
## 87      AM
## 104     AO
## 131     AR
## VISUALIZATIONS :
```
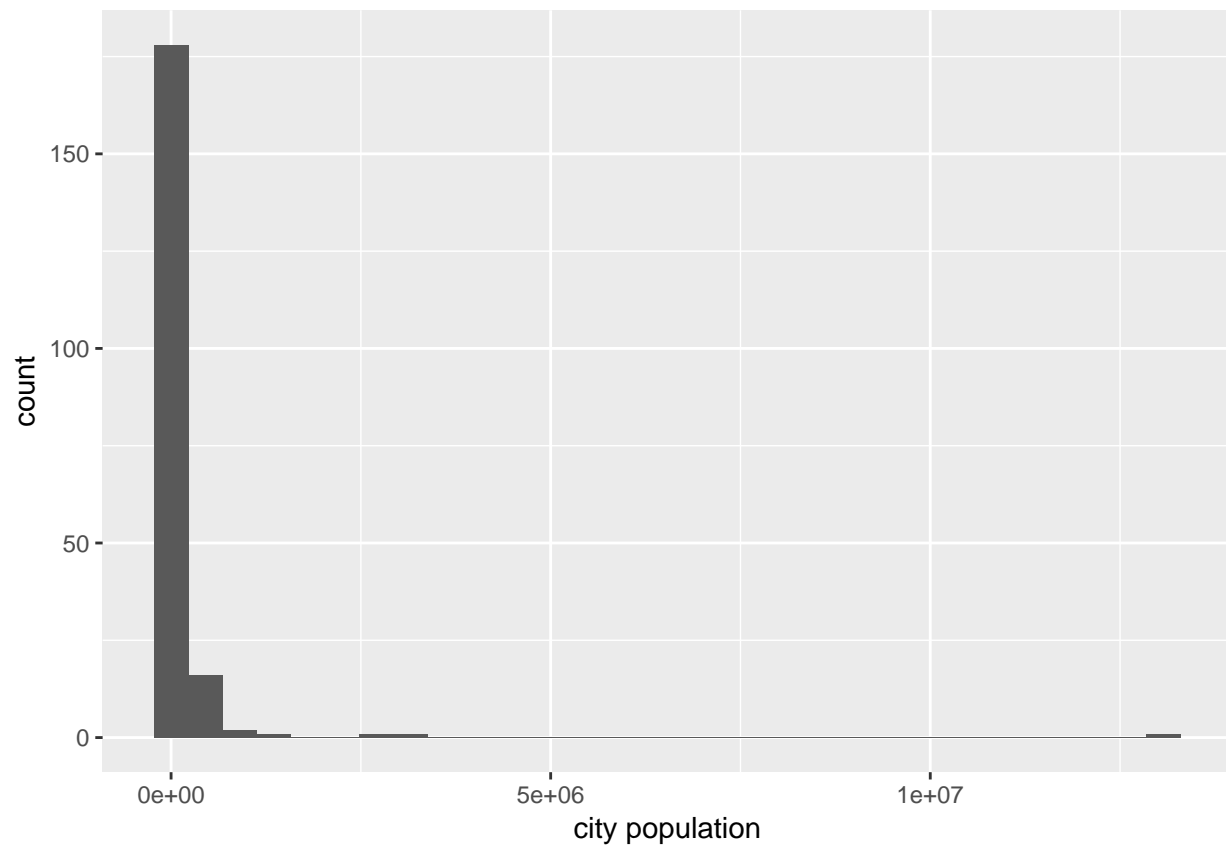
```
WorldCities %>% ggplot(aes(x = population)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
WorldCities %>% ggplot(aes(x = population)) + geom_histogram() + xlab('city population')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
# QUESTION 4 :

WorldCities %>% ggplot(aes(x = population)) + geom_histogram() + xlab('city population') +
ylab('count of people') + ggtitle('My Title') + theme_bw()
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

My Title

count of people

city population