HW_3

Benjamin Ramsey

2/12/2021

The variables I chose to look at in the data set were project cost and the how much incentive (in terms of dollars) the project received. These data were collected by the New York State Energy Research and Development Authority. I chose to focus projects designated as Commercial or Industrial because there are fewer of them. There are some missing values because not all of the data was available. I omitted all the missing values, since they are not useful in any analyses. The data are extremely right-tailed, as shown by the histograms.

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(tinytex)

Solar_Electric_Programs_NYSERDA <- read_csv("Solar_Electric_Programs_NYSERDA.csv")
```

```
##
## -- Column specification ------------------------------------------------------
## cols(
##    .default = col_character(),
##    `Zip Code` = col_double(),
##    `Total Inverter Quantity` = col_double(),
##    `Total PV Module Quantity` = col_double(),
##    `Project Cost` = col_double(),
##    `$Incentive` = col_double(),
##    `Total Nameplate kW DC` = col_double(),
##    `Expected KWh Annual Production` = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
Cost.Incentive <- filter(Solar_Electric_Programs_NYSERDA, `Program Type` == "Commercial/Industrial (Comp
```

```
summarize(Cost.Incentive, mean(`Project Cost`), median(`Project Cost`), mean(`$Incentive`),  median(`$I
```

```
## # A tibble: 1 x 4
##   `mean(\`Project Co~ `median(\`Project C~ `mean(\`$Incentiv~ `median(\`$Incent~
##                <dbl>                <dbl>             <dbl>             <dbl>
## 1           2378666.             1483835           576346.           334973.
```

```
ggplot(data = Cost.Incentive) +
  geom_histogram(mapping = aes(`Project Cost`)) +
  scale_x_log10()
```

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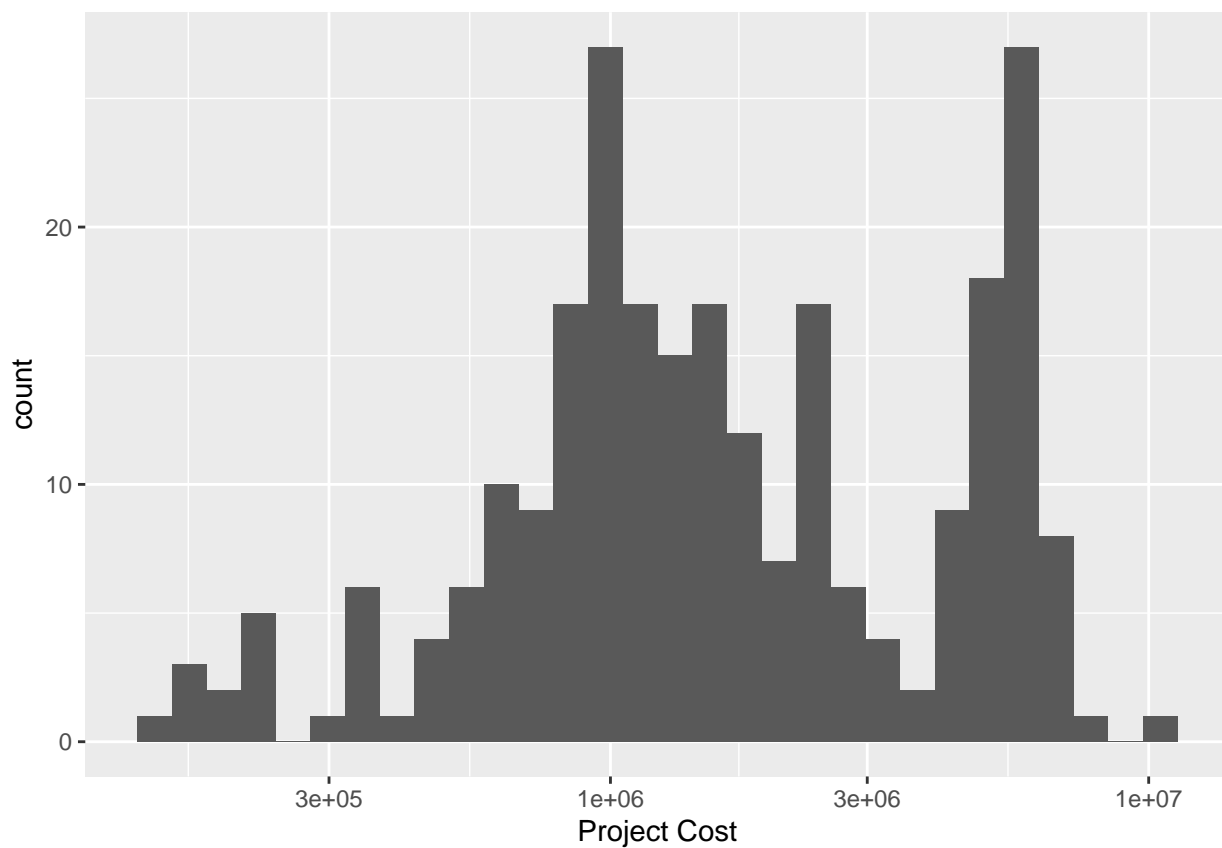

```
ggplot(data = Cost.Incentive) +
  geom_histogram(mapping = aes(`$Incentive`)) +
  scale_x_log10()
```

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