Airline_fatalitiesComparison

Caroline Oliver

4/3/2018

Load in the dataset from csv file

NOTE: Make sure to replace the path variable with your own path to the file

```r
library(readr)
airline_safety <- read_csv("/Users/carolineoliver/airline-safety.csv")
```

```
## Parsed with column specification:
## cols(
##   airline = col_character(),
##   avail_seat_km_per_week = col_double(),
##   incidents_85_99 = col_integer(),
##   fatal_accidents_85_99 = col_integer(),
##   fatalities_85_99 = col_integer(),
##   incidents_00_14 = col_integer(),
##   fatal_accidents_00_14 = col_integer(),
##   fatalities_00_14 = col_integer()
## )
```
```r
# REPLACE LINE ABOVE WITH YOUR PATH: airline_safety <- read_csv("path_to_csv_file_here")
```

Get number of ASK based on trillions (like in the study)

```r
airline_safety$ASKperTrillion = airline_safety[, "avail_seat_km_per_week"]/1000000000000
```

Get number of ASK based on billiosns to compare charts

```r
airline_safety$ASKperBillion = airline_safety[, "avail_seat_km_per_week"]/1000000000
```

add columns to dataset for 85 to 99 (per trillion and per billion)

```r
airline_safety$Fatalities_85_99_trillion = airline_safety$fatalities_85_99 / airline_safety$ASKperTrilli
airline_safety$Fatalities_85_99_billion = airline_safety$fatalities_85_99 / airline_safety$ASKperBillio
```

**add columns to dataset for 00 to 14 (per trillion and per billion)**

```r
airline_safety$Fatalities_00_14_trillion = airline_safety$fatalities_00_14 / airline_safety$ASKperTrilli
airline_safety$Fatalities_00_14_billion = airline_safety$fatalities_00_14 / airline_safety$ASKperBillio
```

Make the plot for 'per trillion'

the data for the columns needed is originally stored as a list

must be converted to numeric vector to be used in plot function

```
tril_85_99 = as.numeric(as.character(unlist(airline_safety$Fatalities_85_99_trillion)))
tril_00_14 = as.numeric(as.character(unlist(airline_safety$Fatalities_00_14_trillion)))
plot(tril_85_99, tril_00_14, xlab="1985-99", ylab="2000-14", main = "Fatalities per Trillion")
abline(lm(tril_85_99 ~ tril_00_14))
```
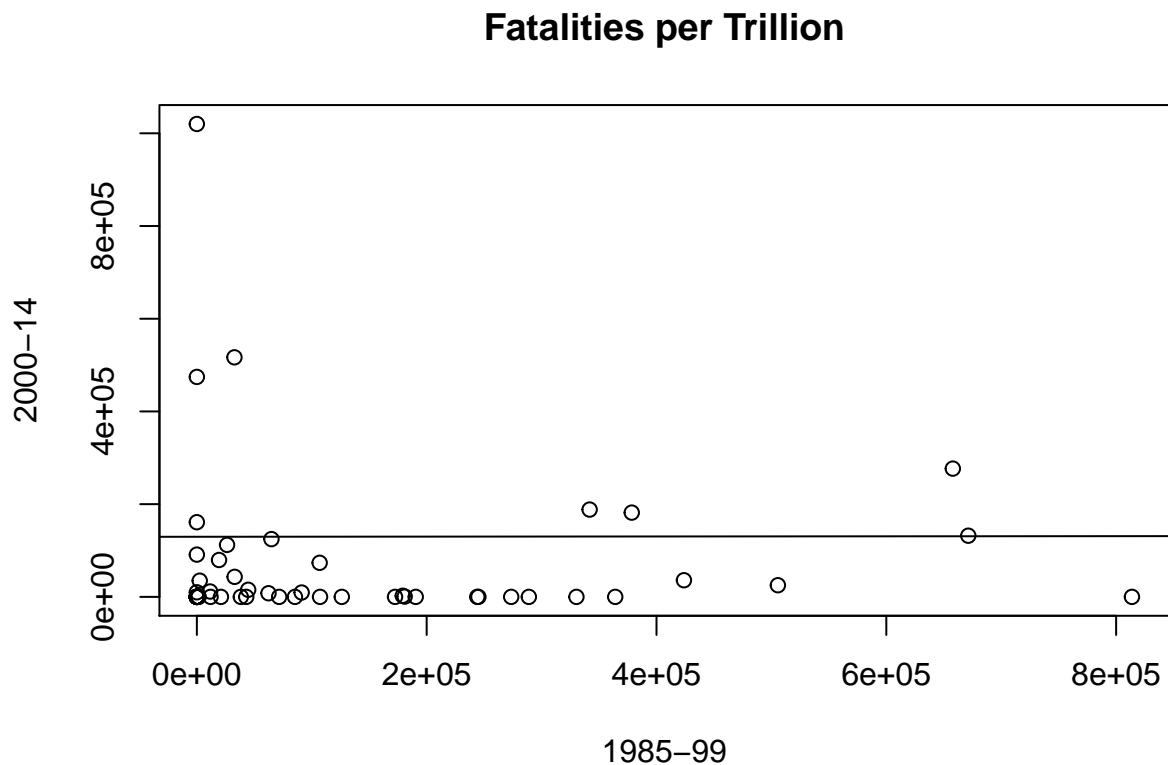

Make the plot for 'per billion'

the data for the columns needed is originally stored as a list

must be converted to numeric vector to be used in plot function

```
bil_85_99 = as.numeric(as.character(unlist(airline_safety$Fatalities_85_99_billion)))
bil_00_14 = as.numeric(as.character(unlist(airline_safety$Fatalities_00_14_billion)))
plot(bil_85_99, bil_00_14, xlab="1985-99", ylab="2000-14", main = "Fatalities per Billion")
abline(lm(bil_85_99 ~ bil_00_14))
```

Fatalities per Billion
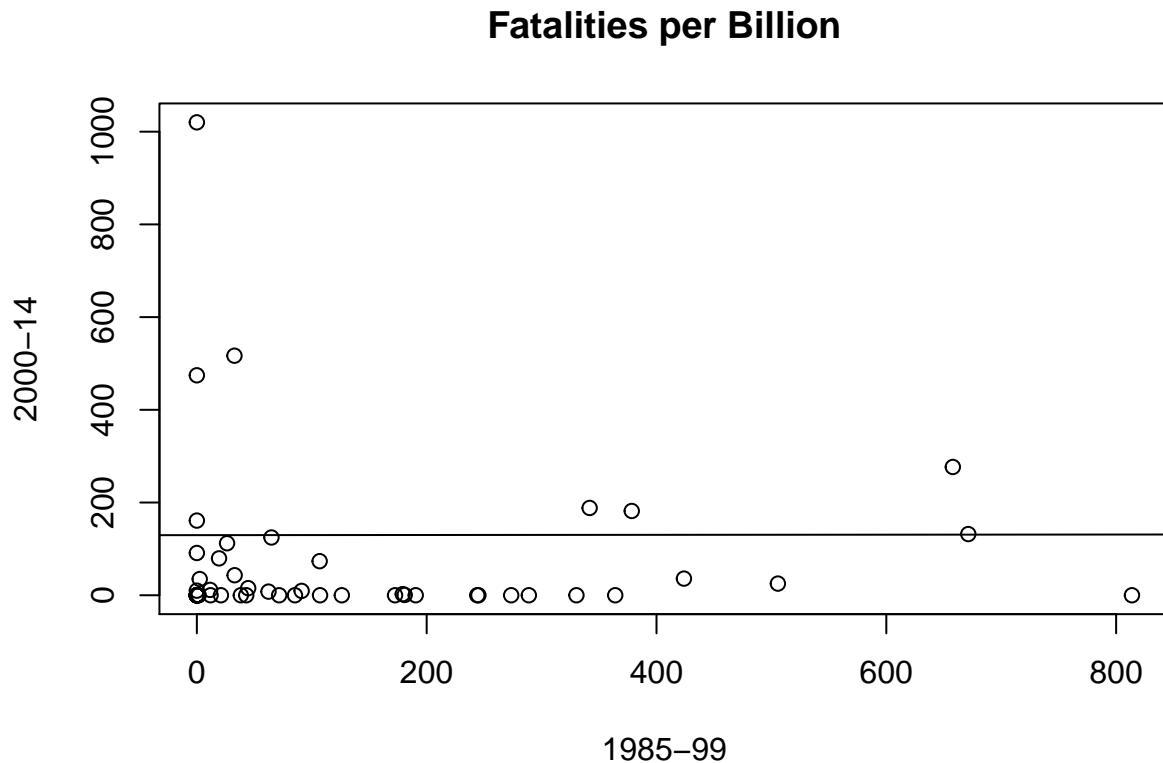
NOTES SO FAR ON GRAPHS ABOVE VS STUDY

Per trillion does not yeild the same result axies as the study

however, the graph points look the same pattern wise

To attain the number of fatalities per trillion, I:

1) Divided total ASK given in first column by a trillion

2) Divided the number of fatalites by the numbers calculated in 1

The same steps were followed to attain per billion numbers

Both scatterplots show no correlation between the two year groups

This is consistent with the study performed

As stated above, all graphs appear to have the points pattern

The only difference appears to be the axes