DATA607 _Assignment5_Rajan

Krishna Rajan

3/4/2018

```
##The chart above describes arrival delays for two airlines across five destinations. Your task is to:
##(1) Create a .CSV file (or optionally, a MySQL database!) that includes all of the information above.
##(2)Readtheinformationfromyour.CSVfileintoR,andusetidyrand dplyr asneededtotidy and transform your data
##(3) Perform analysis to compare the arrival delays for the two airlines.

library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stringr)

flights_data <- read.csv("/Users/rajans/Desktop/CUNY/Data Acquition & Management/Flights_assignment5.csv
head(flights_data)
```

```
##         X     X.1 Los.Angeles Phoenix San.Diego San.Francisco Seattle
## 1  Alaska  Ontime         497     221       212           503    1841
## 2    <NA> Delayed          62      12        20           102     305
## 3    <NA>    <NA>          NA      NA        NA            NA      NA
## 4 AM West  Ontime         694    4840       383           320     201
## 5    <NA> Delayed         117     415        65           129      61
```

```
##Add Column names to X & X1
colnames(flights_data)[1] <- "Airline"
colnames(flights_data)[2] <- "Status"
flights_data[2, 1] <- "Alaska"
flights_data[5, 1] <- "AM West"
head(flights_data)
```

```
##   Airline  Status Los.Angeles Phoenix San.Diego San.Francisco Seattle
## 1  Alaska  Ontime         497     221       212           503    1841
## 2  Alaska Delayed          62      12        20           102     305
## 3    <NA>    <NA>          NA      NA        NA            NA      NA
## 4 AM West  Ontime         694    4840       383           320     201
## 5 AM West Delayed         117     415        65           129      61
```

```
##fib- "fill in the blanks""
##Remove line with NA & add airline names to appropriate rows
flights_data_fib <-na.omit(flights_data)
head(flights_data_fib)
```

```
##   Airline  Status Los.Angeles Phoenix San.Diego San.Francisco Seattle
```

```
## 1  Alaska  Ontime         497     221       212           503    1841
## 2  Alaska Delayed          62      12        20           102     305
## 4 AM West  Ontime         694    4840       383           320     201
## 5 AM West Delayed         117     415        65           129      61
```

```
## T2C: Text to columns
flights_data_fibT2C <- gather(flights_data_fib, "City", "Count", 3:7)
flights_data_fibT2C$City <- str_replace(flights_data_fibT2C$City, "[.]", " ")
flights_data_final <- spread(flights_data_fibT2C, Status, Count)
flights_data_final
```

```
##    Airline          City Delayed Ontime
## 1   Alaska   Los Angeles      62    497
## 2   Alaska       Phoenix      12    221
## 3   Alaska     San Diego      20    212
## 4   Alaska San Francisco     102    503
## 5   Alaska       Seattle     305   1841
## 6  AM West   Los Angeles     117    694
## 7  AM West       Phoenix     415   4840
## 8  AM West     San Diego      65    383
## 9  AM West San Francisco     129    320
## 10 AM West       Seattle      61    201
```

```
##Calculate Delaypercentages between the 2 airlines
DelayOverall <- flights_data_final %>%
  group_by(Airline) %>%
  summarise(TotalDelayed=sum(Delayed),TotalOnTime=sum(`Ontime`), PercentDelayed=round((TotalDelayed/(To

##Caluclate delay by airline & airport
DelayOverall
```

```
## # A tibble: 2 x 4
##   Airline TotalDelayed TotalOnTime PercentDelayed
##   <fct>          <int>       <int>          <dbl>
## 1 Alaska           501        3274           13.3
## 2 AM West          787        6438           10.9
```

```
## AM West has more ontime flights 89.11% as compared to Alaska seems to have more intime flights 86.73%

DelayCity <- flights_data_final %>%
  group_by(Airline, City) %>%
  summarise(TotalDelayed=sum(Delayed),TotalOnTime=sum(`Ontime`), PercentDelayed=round((TotalDelayed/(To

DelayCity
```

```
## # A tibble: 10 x 5
## # Groups:   Airline [?]
##    Airline City          TotalDelayed TotalOnTime PercentDelayed
##    <fct>   <chr>                <int>       <int>          <dbl>
##  1 Alaska  Los Angeles             62         497          11.1
##  2 Alaska  Phoenix                 12         221           5.15
##  3 Alaska  San Diego               20         212           8.62
##  4 Alaska  San Francisco          102         503          16.9
##  5 Alaska  Seattle                305        1841          14.2
##  6 AM West Los Angeles            117         694          14.4
##  7 AM West Phoenix                415        4840           7.90
```

```
##  8 AM West San Diego               65         383          14.5
##  9 AM West San Francisco          129         320          28.7
## 10 AM West Seattle                 61         201          23.3
```

```
## looking at the % delayed by city the data is much clear the bulk of AM West's vast improvement in on
##removing Phoenix from the mix Alaska is more efficient in arrivals as comapred to AM West.
```