

Assignment 5

Subhalaxmi Rout

03/01/2020

		Los Angeles	Phoenix	San Diego	San Francisco	Seattle
ALASKA	on time	497	221	212	503	1,841
	delayed	62	12	20	102	305
AM WEST	on time	694	4,840	383	320	201
	delayed	117	415	65	129	61

Source: [Numbersense](#), Kaiser Fung, McGraw Hill, 2013

Overview

Given above table we can see different Airlines, City and the count of flight ontime and delayed. Analysis below shows, total number of flight delays between Airlines, Total number of flight delays between cities and the delay ratio between different city.

[Note: Flight Delay percent = (Number of flight delayed / Total number of flight) * 100

Flight ontime percent = (Number of flight ontime / Total number of flight) * 100]

Loaded table in 2 ways.

1. Using “insert_table” method
2. After loading table, export table data to csv file and read that file using read.csv().

Installed required packages and loaded libraries

```
#devtools::install_github("lbusett/insert_table")
#install.packages("tidyr")
#install.packages("dplyr")
#install.packages("ggplot2")
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(ggplot2)
library(knitr)
library(inserttable)
```

Crear Table using inserttable method

```
#insert_table(tbl_name = "Flights", nrows = 4, ncols = 4, tbl_format = "DT")
Flights <- tibble::tribble(
  ~AirLines, ~Status, ~LosAngeles, ~Phoenix, ~SanDeigo, ~SanFrancisco, ~Seattle,
  "ALASKA", "ontime", "497", "221", "212", "503", "1841",
  NA, "delayed", "62", "12", "20", "102", "305",
  NA, NA, NA, NA, NA, NA, NA,
  "AMWEST", "ontime", "694", "4840", "383", "320", "201",
  NA, "delayed", "117", "415", "65", "129", "61"
)

require(knitr)
kable(Flights, digits = 3, row.names = FALSE, align = "c",
      caption = NULL)
```

AirLines	Status	LosAngeles	Phoenix	SanDeigo	SanFrancisco	Seattle
ALASKA	ontime	497	221	212	503	1841
NA	delayed	62	12	20	102	305
NA	NA	NA	NA	NA	NA	NA
AMWEST	ontime	694	4840	383	320	201
NA	delayed	117	415	65	129	61

```
Flights <- as.data.frame(Flights)
Flights
```

```
##   AirLines Status LosAngeles Phoenix SanDeigo SanFrancisco 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  AMWEST  ontime      694     4840      383          320     201
## 5   <NA> delayed      117      415       65          129      61
```

```
#Export to csv file
write.csv(Flights, file = "FlightsTable.csv")

#Read csv file
```

```
Flights <- read.csv("FlightsTable.csv")
Flights
```

```
##      X AirLines  Status LosAngeles Phoenix SanDeigo SanFrancisco Seattle
## 1 1    ALASKA  ontime      497      221      212          503      1841
## 2 2      <NA> delayed       62       12       20          102       305
## 3 3      <NA>    <NA>      NA      NA      NA          NA       NA
## 4 4    AMWEST  ontime      694     4840     383          320       201
## 5 5      <NA> delayed      117     415      65          129        61
```

```
#Replace 'NA' with correspond Airlines
```

```
Flights$AirLines[2] <- c("ALASKA")
```

```
Flights$AirLines[5] <- c("AMWEST")
```

```
# Remove NA row from the data
```

```
Flights <- Flights %>% filter(!is.na(Flights$Status))
```

```
Flights
```

```
##      X AirLines  Status LosAngeles Phoenix SanDeigo SanFrancisco Seattle
## 1 1    ALASKA  ontime      497      221      212          503      1841
## 2 2    ALASKA delayed       62       12       20          102       305
## 3 4    AMWEST  ontime      694     4840     383          320       201
## 4 5    AMWEST delayed      117     415      65          129        61
```

```
#Convert differnt city name to column "City"
```

```
Flights_new <- Flights %>% select(AirLines,Status,LosAngeles,Phoenix,SanDeigo,SanFrancisco,Seattle) %>%
```

```
Flights_new
```

```
##      AirLines  Status      City Flight_Count
## 1    ALASKA  ontime  LosAngeles      497
## 2    ALASKA delayed  LosAngeles       62
## 3    AMWEST  ontime  LosAngeles      694
## 4    AMWEST delayed  LosAngeles      117
## 5    ALASKA  ontime   Phoenix      221
## 6    ALASKA delayed   Phoenix       12
## 7    AMWEST  ontime   Phoenix     4840
## 8    AMWEST delayed   Phoenix      415
## 9    ALASKA  ontime   SanDeigo      212
## 10   ALASKA delayed   SanDeigo       20
## 11   AMWEST  ontime   SanDeigo      383
## 12   AMWEST delayed   SanDeigo       65
## 13   ALASKA  ontime SanFrancisco      503
## 14   ALASKA delayed SanFrancisco      102
## 15   AMWEST  ontime SanFrancisco      320
## 16   AMWEST delayed SanFrancisco      129
## 17   ALASKA  ontime   Seattle     1841
## 18   ALASKA delayed   Seattle      305
## 19   AMWEST  ontime   Seattle      201
## 20   AMWEST delayed   Seattle       61
```

```
#Convert "Status" Column to row
Flights_new <- Flights_new %>% mutate_if(is.factor, as.character)
#Flights_new <- tibble::rowid_to_column(Flights_new)
Flights_new
```

```
##      AirLines Status      City Flight_Count
## 1    ALASKA ontime  LosAngeles      497
## 2    ALASKA delayed LosAngeles       62
## 3    AMWEST ontime  LosAngeles     694
## 4    AMWEST delayed LosAngeles     117
## 5    ALASKA ontime   Phoenix      221
## 6    ALASKA delayed Phoenix        12
## 7    AMWEST ontime   Phoenix    4840
## 8    AMWEST delayed Phoenix      415
## 9    ALASKA ontime  SanDeigo     212
## 10   ALASKA delayed SanDeigo       20
## 11   AMWEST ontime  SanDeigo     383
## 12   AMWEST delayed SanDeigo       65
## 13   ALASKA ontime SanFrancisco    503
## 14   ALASKA delayed SanFrancisco   102
## 15   AMWEST ontime SanFrancisco   320
## 16   AMWEST delayed SanFrancisco   129
## 17   ALASKA ontime   Seattle    1841
## 18   ALASKA delayed Seattle      305
## 19   AMWEST ontime   Seattle     201
## 20   AMWEST delayed Seattle       61
```

```
Flights_new <- Flights_new %>% spread(Status,Flight_Count)
Flights_new
```

```
##      AirLines      City delayed ontime
## 1    ALASKA  LosAngeles      62  497
## 2    ALASKA   Phoenix      12  221
## 3    ALASKA  SanDeigo       20  212
## 4    ALASKA SanFrancisco    102  503
## 5    ALASKA   Seattle     305 1841
## 6    AMWEST  LosAngeles     117  694
## 7    AMWEST   Phoenix     415 4840
## 8    AMWEST  SanDeigo       65  383
## 9    AMWEST SanFrancisco    129  320
## 10   AMWEST   Seattle       61  201
```

```
#Analysis: graph shows Different Airline delayed flight count
delay_airline <- Flights_new %>% select(AirLines,delayed,ontime)
delay_airline <- delay_airline %>% group_by(AirLines) %>% summarise(Total_delay_Flight = sum(delayed))
delay_airline
```

```
## # A tibble: 2 x 2
##   AirLines Total_delay_Flight
##   <chr>      <int>
## 1 ALASKA      501
## 2 AMWEST     787
```

```
ggplot(delay_airline, aes(x = delay_airline$AirLines, y = delay_airline$Total_delay_Flight)) + geom_bar
```

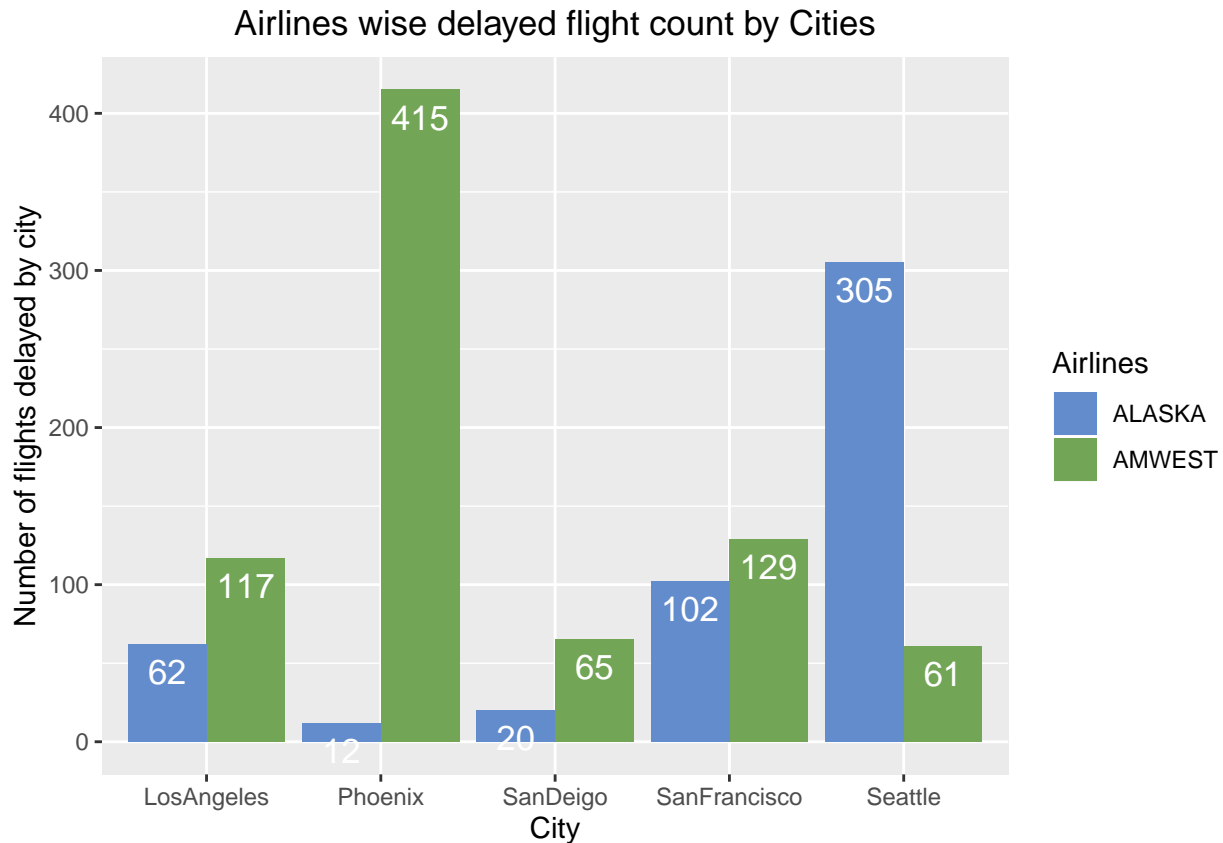


Analysis: Different cities delayed flight count

```
delay_city_flight <- Flights_new %>% select(AirLines, City, delayed)
delay_city_flight <- delay_city_flight %>% group_by(AirLines, City) %>% summarise(Total_delay_Flight = sum(delayed))
delay_city_flight
```

```
## # A tibble: 10 x 3
## # Groups:   AirLines [2]
##   AirLines City      Total_delay_Flight
##   <chr>    <chr>          <int>
## 1 ALASKA  LosAngeles         62
## 2 ALASKA  Phoenix            12
## 3 ALASKA  SanDeigo           20
## 4 ALASKA  SanFrancisco       102
## 5 ALASKA  Seattle            305
## 6 AMWEST  LosAngeles         117
## 7 AMWEST  Phoenix            415
## 8 AMWEST  SanDeigo            65
## 9 AMWEST  SanFrancisco       129
## 10 AMWEST Seattle             61
```

```
ggplot(data = delay_city_flight, aes(x = delay_city_flight$City, y = delay_city_flight$Total_delay_Flight))
```

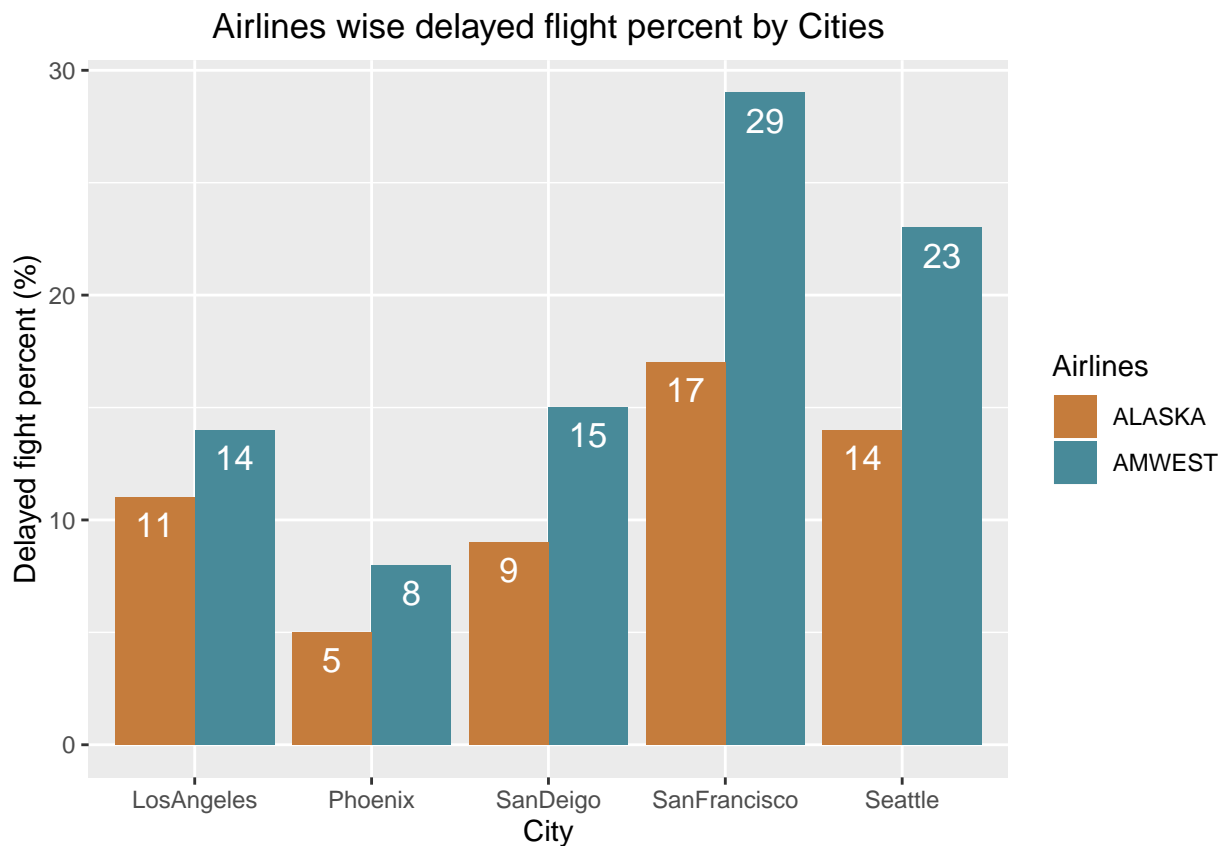


Analysis: Flight Delay percent by different cities

```
delay_percent <- Flights_new %>% select(AirLines, City, ontime, delayed) %>% group_by(AirLines, City) %>% mutate(percent = as.integer(round(delayed / total_flight * 100)))
delay_percent
```

```
## # A tibble: 10 x 6
## # Groups:   AirLines, City [10]
##   AirLines City      ontime delayed total_flight percent
##   <chr>    <chr>      <int>   <int>      <int>    <int>
## 1 ALASKA  LosAngeles    497     62       559      11
## 2 ALASKA  Phoenix       221     12       233      5
## 3 ALASKA  SanDeigo      212     20       232      9
## 4 ALASKA  SanFrancisco  503    102       605     17
## 5 ALASKA  Seattle     1841   305      2146     14
## 6 AMWEST  LosAngeles    694    117       811     14
## 7 AMWEST  Phoenix     4840   415      5255      8
## 8 AMWEST  SanDeigo     383     65       448     15
## 9 AMWEST  SanFrancisco  320    129       449     29
## 10 AMWEST Seattle       201     61       262     23
```

```
ggplot(data = delay_percent, aes(x = delay_percent$City, y = delay_percent$percent, fill = delay_percent$Airline))
```

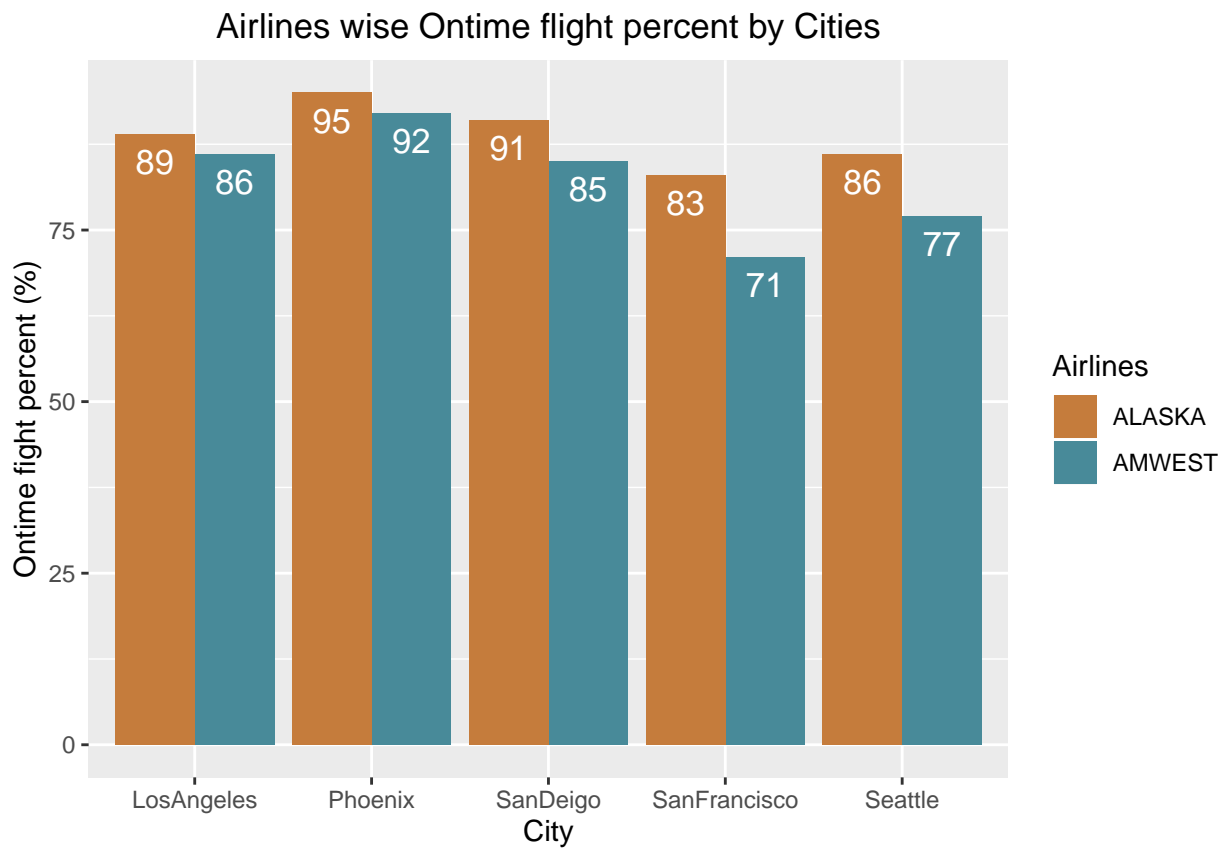


Analysis: Flight ontime percent by different cities

```
ontime_percent <- Flights_new %>% select(AirLines, City, ontime, delayed) %>% group_by(AirLines, City) %>% summarise(
  ontime_percent = ontime_percent %>% group_by(AirLines, City) %>% mutate(percent_ontime = as.integer(round(
  ontime_percent
```

```
## # A tibble: 10 x 6
## # Groups:   AirLines, City [10]
##   AirLines City      ontime delayed total_flight percent_ontime
##   <chr>    <chr>      <int>   <int>      <int>         <int>
## 1 ALASKA  LosAngeles    497     62       559           89
## 2 ALASKA  Phoenix       221     12       233           95
## 3 ALASKA  SanDeigo      212     20       232           91
## 4 ALASKA  SanFrancisco  503    102       605           83
## 5 ALASKA  Seattle     1841    305     2146           86
## 6 AMWEST  LosAngeles    694    117       811           86
## 7 AMWEST  Phoenix     4840    415     5255           92
## 8 AMWEST  SanDeigo     383     65       448           85
## 9 AMWEST  SanFrancisco  320    129       449           71
## 10 AMWEST Seattle       201     61       262           77
```

```
ggplot(data = ontime_percent, aes(x = ontime_percent$City, y = ontime_percent$percent_ontime, fill = de
```



Conclusion

From the above graph we can see,

- Phoenix city has more ontime flight percent and less delayed flight percent.
- San Francisco has less ontime percent and more delayed percent.
- AMWEST has more delay flights than Alaska.