MY-DATA607-Week05-Flights

Michael Y.

September 30, 2018

Week 5 - Flight Delays

The assignment is as follows:

A chart is supplied, listing flight performance (on-time vs. delayed) for two airlines (ALASKA and AM WEST) across five cities (Los Angeles, Phoenix, San Diego, San Francisco, and Seattle).

(1) Create a .CSV file (or optionally, a MySQL database!) that includes all of the information above. You're encouraged to use a "wide" structure similar to how the information appears above, so that you can practice tidying and transformations as described below.

(2) Read the information from your .CSV file into R, and use `tidyr` and `dplyr` as needed to tidy and transform your data.

(3) Perform analysis to compare the arrival delays for the two airlines.

(4) Your code should be in an R Markdown file, posted to rpubs.com, and should include narrative descriptions of your data cleanup work, analysis, and conclusions.

```
#library(readr)
#library(stringr)
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)
```

Data Loading

Load the raw datafile (which I created by entering the given values into an excel spreadsheet, then saving the spreadsheet as a .csv file) :

```r
##setwd("C:/Users/Michael/Dropbox/priv/CUNY/MSDS/201809-Fall/DATA607_Andy_Sabrina/Week05")
##inputfile <- "InputFlightData.csv"
inputfile <- "https://raw.githubusercontent.com/myampol/MY607/master/InputFlightData.csv"
rawflights <- read.csv(inputfile,stringsAsFactors = F)
head(rawflights)
```

```
##         X     X.1 Los.Angeles Phoenix San.Diego San.Francisco Seattle
## 1  ALASKA on time         497     221       212           503    1841
## 2         delayed          62      12        20           102     305
## 3                          NA      NA        NA            NA      NA
## 4 AM WEST on time         694    4840       383           320     201
## 5         delayed         117     415        65           129      61
```

```r
rf1 <- rename(.data = rawflights, Airline=X, Status=`X.1`,
              LosAngeles=`Los.Angeles`, SanDiego=`San.Diego`, SanFrancisco=`San.Francisco`)
rf1
```

```
##   Airline  Status LosAngeles Phoenix SanDiego SanFrancisco Seattle
## 1  ALASKA on time        497     221      212          503    1841
## 2         delayed         62      12       20          102     305
## 3                         NA      NA       NA           NA      NA
## 4 AM WEST on time        694    4840      383          320     201
## 5         delayed        117     415       65          129      61
```

Data Cleanup Work

Reviewing Hadley Wickham's directive for creating "tidy" data, it is necessary to distinguish between *Fixed Variables* and *Measured Variables* .

Here the *Fixed Variables* include

(a) the airline (AM West or Alaska), and

(b) the city (Los Angeles, Phoenix, San Diego, San Francisco, and Seattle), while

the *Measured Variables* include

(c) the count of ON TIME flights and

(d) the count of DELAYED flights

for each (Airline,City) pair.

First, use gather to put all the Cities into one column, and also drop the row containing NAs:

```r
rf2 <- gather(data = rf1 , key = City, value = NumFlights, ... = LosAngeles:Seattle, na.rm = T)
rf2
```

```
##    Airline  Status         City NumFlights
## 1   ALASKA on time   LosAngeles        497
## 2          delayed   LosAngeles         62
## 4  AM WEST on time   LosAngeles        694
## 5          delayed   LosAngeles        117
## 6   ALASKA on time      Phoenix        221
## 7          delayed      Phoenix         12
## 9  AM WEST on time      Phoenix       4840
## 10         delayed      Phoenix        415
## 11  ALASKA on time     SanDiego        212
## 12         delayed     SanDiego         20
## 14 AM WEST on time     SanDiego        383
## 15         delayed     SanDiego         65
## 16  ALASKA on time SanFrancisco        503
## 17         delayed SanFrancisco        102
## 19 AM WEST on time SanFrancisco        320
## 20         delayed SanFrancisco        129
## 21  ALASKA on time      Seattle       1841
## 22         delayed      Seattle        305
## 24 AM WEST on time      Seattle        201
## 25         delayed      Seattle         61
```

Next, use mutate(lag) to propagate the Airline names downward (from each odd-numbered row) to fill the missing airline name (on the subsequent even-numbered row):

```r
rf3 <- mutate(.data = rf2, Airline= ifelse(Airline=="", lag(Airline), Airline))
rf3
```

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

Now, use spread to put the "on time" and "delayed" counts into separate columns:

```r
rf4 <- spread(data = rf3, key = Status, value = NumFlights)
rf4
```

```
##    Airline         City delayed on time
## 1   ALASKA   LosAngeles      62     497
## 2   ALASKA      Phoenix      12     221
## 3   ALASKA     SanDiego      20     212
## 4   ALASKA SanFrancisco     102     503
## 5   ALASKA      Seattle     305    1841
## 6  AM WEST   LosAngeles     117     694
## 7  AM WEST      Phoenix     415    4840
## 8  AM WEST     SanDiego      65     383
## 9  AM WEST SanFrancisco     129     320
## 10 AM WEST      Seattle      61     201
```

Use rename to improve the names of the "on time" and "delayed" columns:

```
rf5 <- rename(.data = rf4, NumFlightsDelayed=delayed, NumFlightsOnTime=`on time`)
rf5
```

```
##    Airline         City NumFlightsDelayed NumFlightsOnTime
## 1   ALASKA   LosAngeles                62              497
## 2   ALASKA      Phoenix                12              221
## 3   ALASKA     SanDiego                20              212
## 4   ALASKA SanFrancisco               102              503
## 5   ALASKA      Seattle               305             1841
## 6  AM WEST   LosAngeles               117              694
## 7  AM WEST      Phoenix               415             4840
## 8  AM WEST     SanDiego                65              383
## 9  AM WEST SanFrancisco               129              320
## 10 AM WEST      Seattle                61              201
```

Use mutate to compute and append NumFlightsTotal:

```
rf6 <- mutate(.data = rf5, NumFlightsTotal = NumFlightsDelayed + NumFlightsOnTime)
rf6
```

```
##    Airline         City NumFlightsDelayed NumFlightsOnTime NumFlightsTotal
## 1   ALASKA   LosAngeles                62              497             559
## 2   ALASKA      Phoenix                12              221             233
## 3   ALASKA     SanDiego                20              212             232
## 4   ALASKA SanFrancisco               102              503             605
## 5   ALASKA      Seattle               305             1841            2146
## 6  AM WEST   LosAngeles               117              694             811
## 7  AM WEST      Phoenix               415             4840            5255
## 8  AM WEST     SanDiego                65              383             448
## 9  AM WEST SanFrancisco               129              320             449
## 10 AM WEST      Seattle                61              201             262
```

Use mutate to compute and append the *percentage* of delayed and ontime flights at each city:

```
rf7 <- mutate(.data = rf6, PctFlightsDelayed=NumFlightsDelayed/NumFlightsTotal,
                           PctFlightsOnTime=NumFlightsOnTime/NumFlightsTotal)
rf7
```

```
##    Airline         City NumFlightsDelayed NumFlightsOnTime NumFlightsTotal
## 1   ALASKA   LosAngeles                62              497             559
```

```
## 2   ALASKA      Phoenix                12             221            233
## 3   ALASKA     SanDiego                20             212            232
## 4   ALASKA SanFrancisco               102             503            605
## 5   ALASKA      Seattle               305            1841           2146
## 6  AM WEST   LosAngeles               117             694            811
## 7  AM WEST      Phoenix               415            4840           5255
## 8  AM WEST     SanDiego                65             383            448
## 9  AM WEST SanFrancisco               129             320            449
## 10 AM WEST      Seattle                61             201            262
##    PctFlightsDelayed PctFlightsOnTime
## 1         0.11091234        0.8890877
## 2         0.05150215        0.9484979
## 3         0.08620690        0.9137931
## 4         0.16859504        0.8314050
## 5         0.14212488        0.8578751
## 6         0.14426634        0.8557337
## 7         0.07897241        0.9210276
## 8         0.14508929        0.8549107
## 9         0.28730512        0.7126949
## 10        0.23282443        0.7671756
```

The above shows the data tidying and manipulation step-by-step.

Using the pipe connector "%>%", all the above steps can be specified in a single chain:

```
tidy_flights <- rawflights %>%
  rename(.data = ., Airline=X, Status=`X.1`, LosAngeles=`Los.Angeles`,
                                    SanDiego=`San.Diego`,
                                    SanFrancisco=`San.Francisco`)                 %>%
  gather( data = ., key = City, value = NumFlights, ... = LosAngeles:Seattle, na.rm = T)   %>%
  mutate(.data = ., Airline= ifelse(Airline=="", lag(Airline), Airline))          %>%
  spread( data = ., key = Status, value = NumFlights)                             %>%
  rename(.data = ., NumFlightsDelayed=delayed, NumFlightsOnTime=`on time`)        %>%
  mutate(.data = ., NumFlightsTotal = NumFlightsDelayed + NumFlightsOnTime)       %>%
  mutate(.data = ., PctFlightsDelayed=NumFlightsDelayed/NumFlightsTotal,
                    PctFlightsOnTime=NumFlightsOnTime/NumFlightsTotal)
tidy_flights
```

```
##    Airline         City NumFlightsDelayed NumFlightsOnTime NumFlightsTotal
## 1   ALASKA   LosAngeles                62              497             559
## 2   ALASKA      Phoenix                12              221             233
## 3   ALASKA     SanDiego                20              212             232
## 4   ALASKA SanFrancisco               102              503             605
## 5   ALASKA      Seattle               305             1841            2146
## 6  AM WEST   LosAngeles               117              694             811
## 7  AM WEST      Phoenix               415             4840            5255
## 8  AM WEST     SanDiego                65              383             448
## 9  AM WEST SanFrancisco               129              320             449
```

```
## 10 AM WEST      Seattle                61             201             262
##    PctFlightsDelayed PctFlightsOnTime
## 1         0.11091234        0.8890877
## 2         0.05150215        0.9484979
## 3         0.08620690        0.9137931
## 4         0.16859504        0.8314050
## 5         0.14212488        0.8578751
## 6         0.14426634        0.8557337
## 7         0.07897241        0.9210276
## 8         0.14508929        0.8549107
## 9         0.28730512        0.7126949
## 10        0.23282443        0.7671756
```

The above result matches that from the step-by-step process.

Analyze the data

Let's sort the above data by city, then by airline:

```
arrange(tidy_flights, City, Airline)
```

```
##    Airline          City NumFlightsDelayed NumFlightsOnTime NumFlightsTotal
## 1   ALASKA   LosAngeles                62              497             559
## 2  AM WEST   LosAngeles               117              694             811
## 3   ALASKA      Phoenix                12              221             233
## 4  AM WEST      Phoenix               415             4840            5255
## 5   ALASKA     SanDiego                20              212             232
## 6  AM WEST     SanDiego                65              383             448
## 7   ALASKA SanFrancisco               102              503             605
## 8  AM WEST SanFrancisco               129              320             449
## 9   ALASKA      Seattle               305             1841            2146
## 10 AM WEST      Seattle                61              201             262
##    PctFlightsDelayed PctFlightsOnTime
## 1         0.11091234        0.8890877
## 2         0.14426634        0.8557337
## 3         0.05150215        0.9484979
## 4         0.07897241        0.9210276
## 5         0.08620690        0.9137931
## 6         0.14508929        0.8549107
## 7         0.16859504        0.8314050
## 8         0.28730512        0.7126949
## 9         0.14212488        0.8578751
## 10        0.23282443        0.7671756
```

```
ALASKA_Phoenix_delays <- filter(.data=tidy_flights, Airline=="ALASKA" & City=="Phoenix") %>%
  select(PctFlightsDelayed)
AMWEST_Phoenix_delays <- filter(.data=tidy_flights, Airline=="AM WEST" & City=="Phoenix") %>%
  select(PctFlightsDelayed)
ALASKA_SanFrancisco_delays <- filter(.data=tidy_flights, Airline=="ALASKA" & City=="SanFrancisco") %>%
  select(PctFlightsDelayed)
AMWEST_SanFrancisco_delays <- filter(.data=tidy_flights, Airline=="AM WEST" & City=="SanFrancisco") %>%
  select(PctFlightsDelayed)
```

What's noticeable is that looking city-by-city, the Percentage of Flights Delayed is smaller for ALASKA than it is for AMWEST.

Percentage of flights delayed for each airline, by city:

```
Pct_Delays_by_City <- tidy_flights %>% select(.data = ., -NumFlightsDelayed, -NumFlightsOnTime, -NumFli
  spread(data = . , key = Airline, value = PctFlightsDelayed)
Pct_Delays_by_City
```

```
##          City    ALASKA   AM WEST
## 1   LosAngeles 0.11091234 0.14426634
## 2      Phoenix 0.05150215 0.07897241
## 3     SanDiego 0.08620690 0.14508929
## 4 SanFrancisco 0.16859504 0.28730512
## 5      Seattle 0.14212488 0.23282443
```

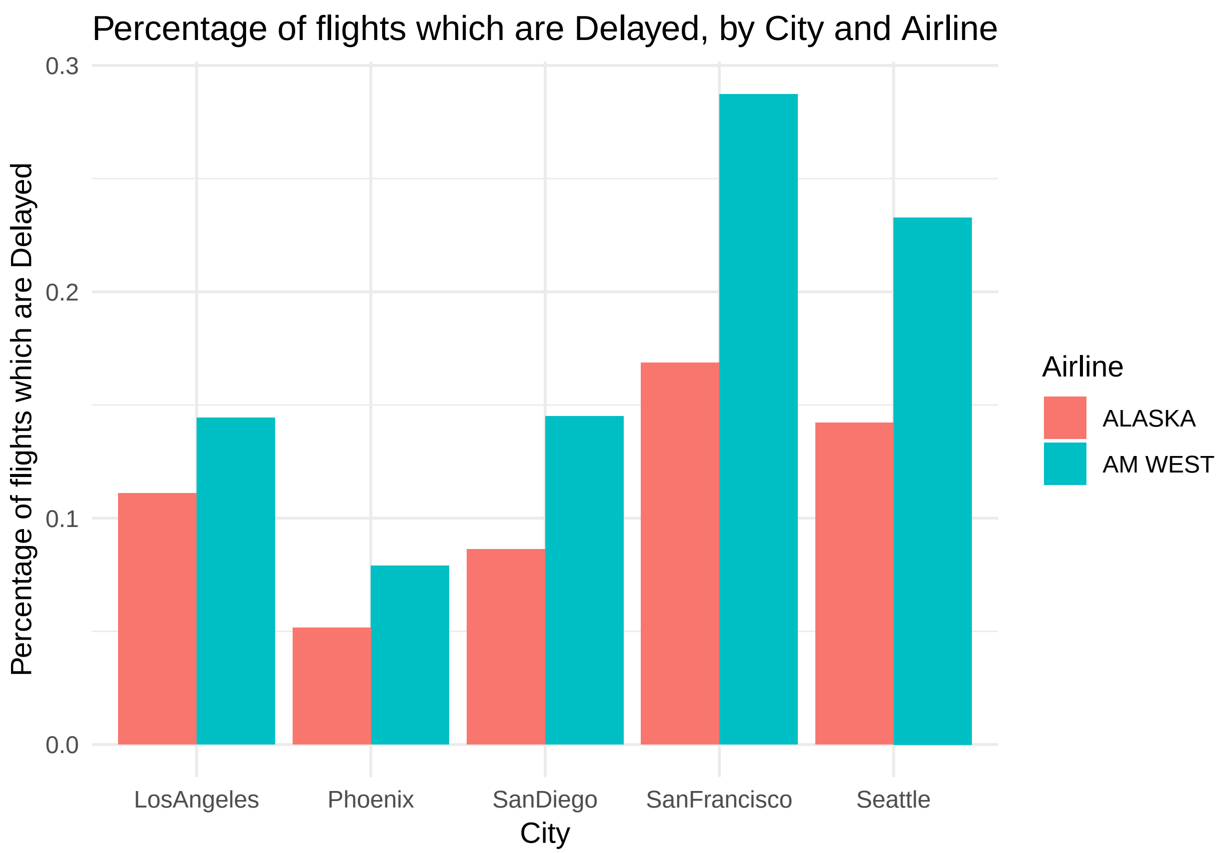
For example, at Phoenix, ALASKA's delays are 5.2% while AM WEST's delays are 7.9% .

At SanFrancisco, ALASKA's delays are 16.9% while AM WEST's delays are 28.7% .

The above relationship holds *for each city.*

Here's a plot, to illustrate:

```
Pct_Delays_by_City %>% gather(data = ., key = Airline, value = Pct_Delays_by_City,...=ALASKA:`AM WEST`)
    ggplot(data = ., aes(factor(City), Pct_Delays_by_City, fill = Airline)) +
    geom_bar(stat="identity", position = "dodge") +
    theme_minimal()+
    labs( x="City", y="Percentage of flights which are Delayed") +
    ggtitle("Percentage of flights which are Delayed, by City and Airline")
```



Question: How many TOTAL flights does each airline have, and what percent are delayed?

```
ResultsByAirline <- group_by(tidy_flights,Airline) %>% summarize(
                                              TotalDelays=sum(NumFlightsDelayed),
                                              TotalOnTime=sum(NumFlightsOnTime),
                                              TotalFlights=sum(NumFlightsTotal),
                                              PctDelayed=TotalDelays/TotalFlights,
                                              PctOnTime=TotalOnTime/TotalFlights)
ResultsByAirline
```

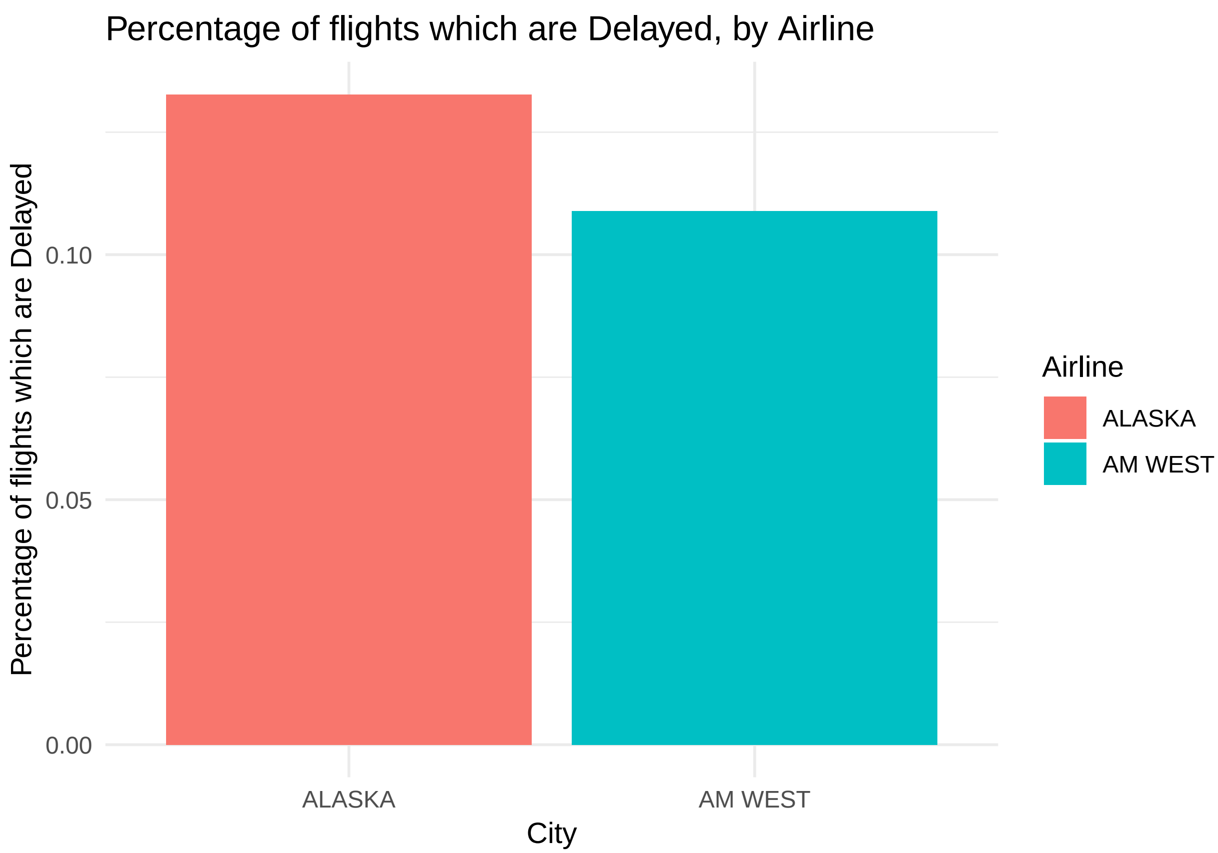
```
## # A tibble: 2 x 6
##   Airline TotalDelays TotalOnTime TotalFlights PctDelayed PctOnTime
##   <chr>         <int>       <int>        <int>      <dbl>     <dbl>
## 1 ALASKA          501        3274         3775      0.133     0.867
## 2 AM WEST         787        6438         7225      0.109     0.891
```

```
ALASKAdelays <- filter(.data = ResultsByAirline, Airline=="ALASKA") %>% select(PctDelayed)
AMWESTdelays <- filter(.data = ResultsByAirline, Airline=="AM WEST") %>% select(PctDelayed)
ALASKAtotals <- filter(.data = ResultsByAirline, Airline=="ALASKA") %>% select(TotalFlights)
AMWESTtotals <- filter(.data = ResultsByAirline, Airline=="AM WEST") %>% select(TotalFlights)
```

These results show that while ALASKA runs about half as many flights (3775) as its competitor AM WEST (7225),

a larger percentage (13.3 %) of ALASKA's flights are delayed vs. AM WEST, which suffered delays on only 10.9 % of its flights.

```
ResultsByAirline %>% select(.data = ., Airline, PctDelayed) %>%
    ggplot(data = ., aes(factor(Airline), PctDelayed, fill = Airline)) +
    geom_bar(stat="identity", position = "dodge") +
    theme_minimal()+
    labs( x="City", y="Percentage of flights which are Delayed") +
    ggtitle("Percentage of flights which are Delayed, by Airline")
```



So, this is curious!

On a city-by-city basis, ALASKA "beat" AM WEST by having better on-time performance at every city.

But on an overall basis, AM WEST had the best overall on-time results!

How could this paradox be explained?

Although the data shows that each airline serves the same 5 cities, they seem to focus on different markets. Perhaps looking more closely at the different cities served by each airline can help explain?

First, let's determine how many TOTAL flights go to each city, and what percent are delayed?

```
ResultsByCity <- group_by(tidy_flights,City) %>% summarize(
                                               TotalDelays=sum(NumFlightsDelayed),
                                               TotalOnTime=sum(NumFlightsOnTime),
                                               TotalFlights=sum(NumFlightsTotal),
                                               PctDelayed=TotalDelays/TotalFlights,
                                               PctOnTime=TotalOnTime/TotalFlights)
ResultsByCity
```

```
## # A tibble: 5 x 6
##   City          TotalDelays TotalOnTime TotalFlights PctDelayed PctOnTime
##   <chr>               <int>       <int>        <int>      <dbl>     <dbl>
## 1 LosAngeles            179        1191         1370     0.131      0.869
## 2 Phoenix               427        5061         5488     0.0778     0.922
## 3 SanDiego               85         595          680     0.125      0.875
## 4 SanFrancisco          231         823         1054     0.219      0.781
## 5 Seattle               366        2042         2408     0.152      0.848
```

```
PhoenixDelays      <- filter(.data = ResultsByCity, City=="Phoenix") %>% select(PctDelayed)
SanFranciscoDelays <- filter(.data = ResultsByCity, City=="SanFrancisco") %>% select(PctDelayed)
```

This shows that the smallest percentage (7.8%) of flights are delayed at Phoenix,

while the largest percentage (21.9%) of flights are delayed at SanFrancisco.

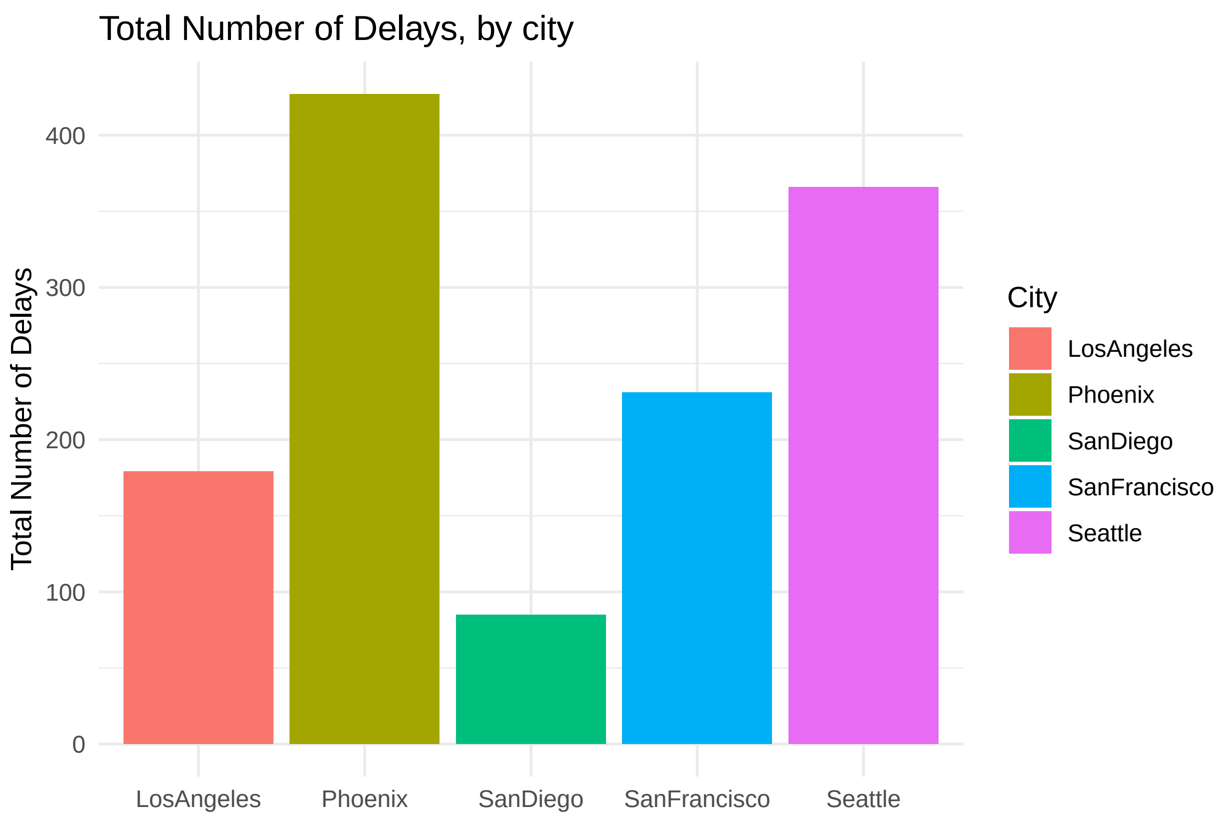
ResultsByCity, sorted by PctDelayed:

```
arrange(ResultsByCity, PctDelayed)
```

```
## # A tibble: 5 x 6
##   City          TotalDelays TotalOnTime TotalFlights PctDelayed PctOnTime
##   <chr>               <int>       <int>        <int>      <dbl>     <dbl>
## 1 Phoenix               427        5061         5488     0.0778     0.922
## 2 SanDiego               85         595          680     0.125      0.875
## 3 LosAngeles            179        1191         1370     0.131      0.869
## 4 Seattle               366        2042         2408     0.152      0.848
## 5 SanFrancisco          231         823         1054     0.219      0.781
```

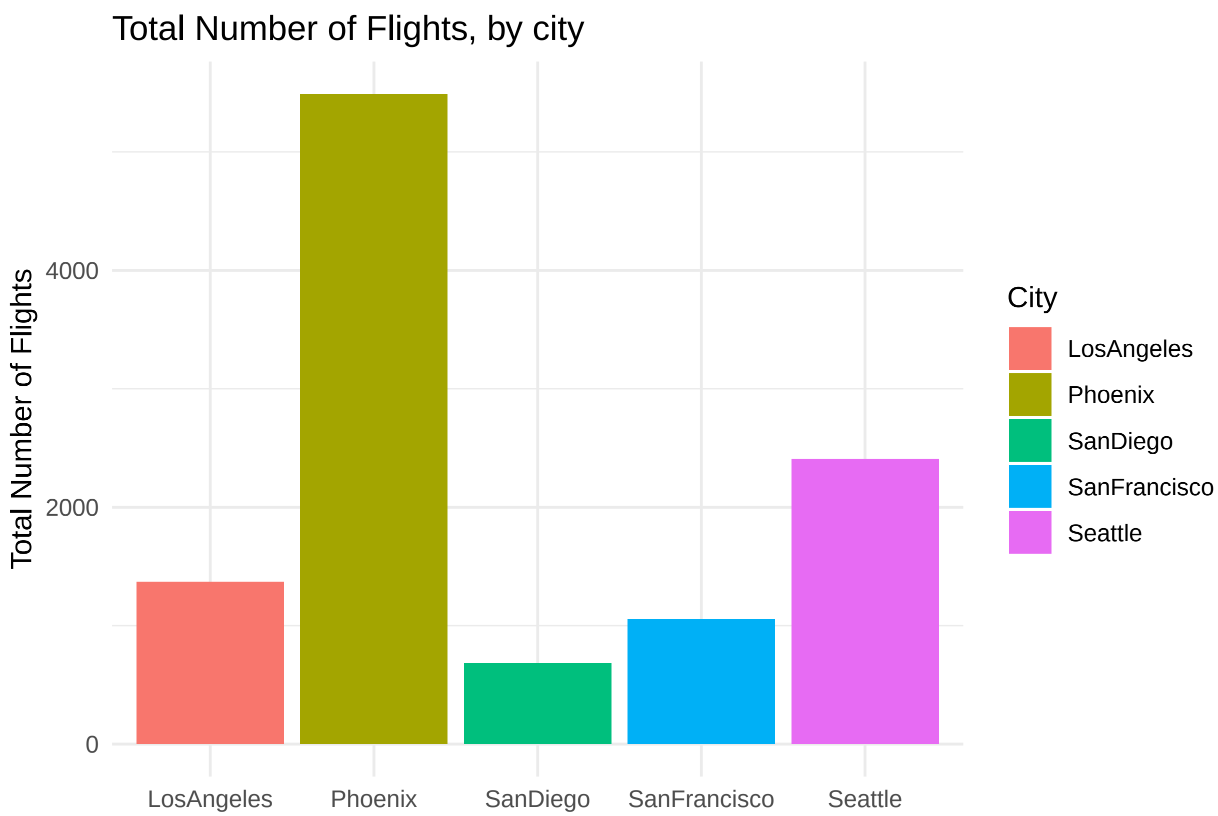
Here's a barplot:

```
select(.data = ResultsByCity, City, TotalDelays) %>%
    ggplot(data = ., aes(factor(City), TotalDelays, fill = City)) +
    geom_bar(stat="identity", position = "dodge") +
    theme_minimal()+
    labs( x="", y="Total Number of Delays") +
    ggtitle("Total Number of Delays, by city")
```



Although the largest absolute number of *delays* occurs in Phoenix, it is the city with the largest number of *overall* flights. Indeed, the percentage of delays at Phoenix is the lowest across all 5 cities.

```
select(.data = ResultsByCity, City, TotalFlights) %>%
    ggplot(data = ., aes(factor(City), TotalFlights, fill = City)) +
    geom_bar(stat="identity", position = "dodge") +
    theme_minimal()+
    labs( x="", y="Total Number of Flights") +
    ggtitle("Total Number of Flights, by city")
```



Let's determine the relative market share of each airline at each city, to see if that helps explain the delays:

Use merge to join the *Totals(by city)* onto the *tidy_flights* dataframe.

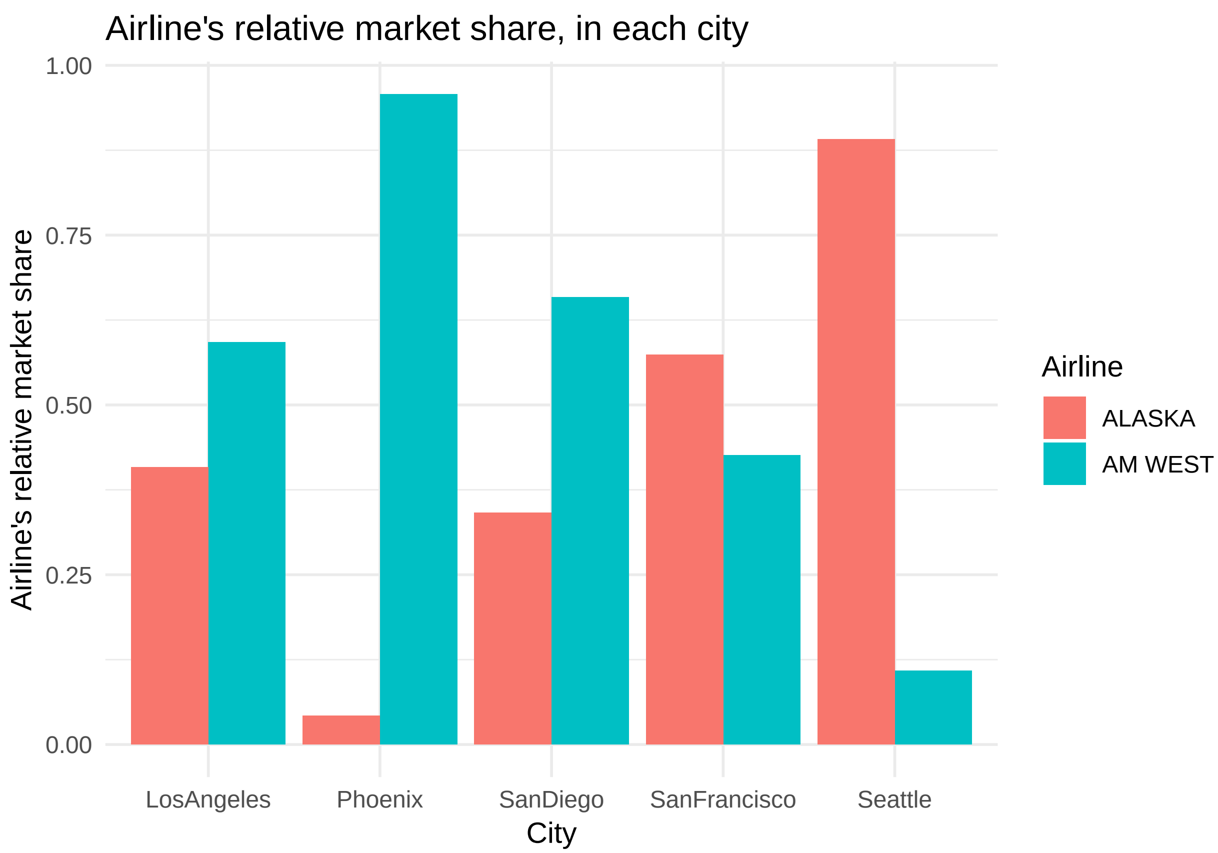
```
big_flights <- tidy_flights %>%
  merge(ResultsByCity) %>%
  arrange(.data = ., City, Airline) %>%
  mutate(.data = ., ShareOfDelays=NumFlightsDelayed/TotalDelays,
                    ShareOfOnTime=NumFlightsOnTime/TotalOnTime,
                    ShareOfFlights=NumFlightsTotal/TotalFlights)
big_flights
```

```
##             City Airline NumFlightsDelayed NumFlightsOnTime NumFlightsTotal
## 1     LosAngeles  ALASKA                62              497             559
## 2     LosAngeles AM WEST               117              694             811
## 3        Phoenix  ALASKA                12              221             233
## 4        Phoenix AM WEST               415             4840            5255
## 5       SanDiego  ALASKA                20              212             232
## 6       SanDiego AM WEST                65              383             448
## 7   SanFrancisco  ALASKA               102              503             605
## 8   SanFrancisco AM WEST               129              320             449
## 9        Seattle  ALASKA               305             1841            2146
## 10       Seattle AM WEST                61              201             262
##    PctFlightsDelayed PctFlightsOnTime TotalDelays TotalOnTime TotalFlights
```

```
## 1          0.11091234         0.8890877          179          1191           1370
## 2          0.14426634         0.8557337          179          1191           1370
## 3          0.05150215         0.9484979          427          5061           5488
## 4          0.07897241         0.9210276          427          5061           5488
## 5          0.08620690         0.9137931           85           595            680
## 6          0.14508929         0.8549107           85           595            680
## 7          0.16859504         0.8314050          231           823           1054
## 8          0.28730512         0.7126949          231           823           1054
## 9          0.14212488         0.8578751          366          2042           2408
## 10         0.23282443         0.7671756          366          2042           2408
##    PctDelayed PctOnTime ShareOfDelays ShareOfOnTime ShareOfFlights
## 1  0.13065693 0.8693431    0.34636872    0.41729639     0.40802920
## 2  0.13065693 0.8693431    0.65363128    0.58270361     0.59197080
## 3  0.07780612 0.9221939    0.02810304    0.04366726     0.04245627
## 4  0.07780612 0.9221939    0.97189696    0.95633274     0.95754373
## 5  0.12500000 0.8750000    0.23529412    0.35630252     0.34117647
## 6  0.12500000 0.8750000    0.76470588    0.64369748     0.65882353
## 7  0.21916509 0.7808349    0.44155844    0.61117861     0.57400380
## 8  0.21916509 0.7808349    0.55844156    0.38882139     0.42599620
## 9  0.15199336 0.8480066    0.83333333    0.90156709     0.89119601
## 10 0.15199336 0.8480066    0.16666667    0.09843291     0.10880399
```

Plot airline market share, by City

```
select(.data = big_flights, City=City,Airline=Airline,ShareOfFlights=ShareOfFlights) %>%
    ggplot(data = ., aes(factor(City), ShareOfFlights, fill = Airline)) +
    geom_bar(stat="identity", position = "dodge") +
    theme_minimal()+
    labs( x="City", y="Airline's relative market share") +
    ggtitle("Airline's relative market share, in each city")
```



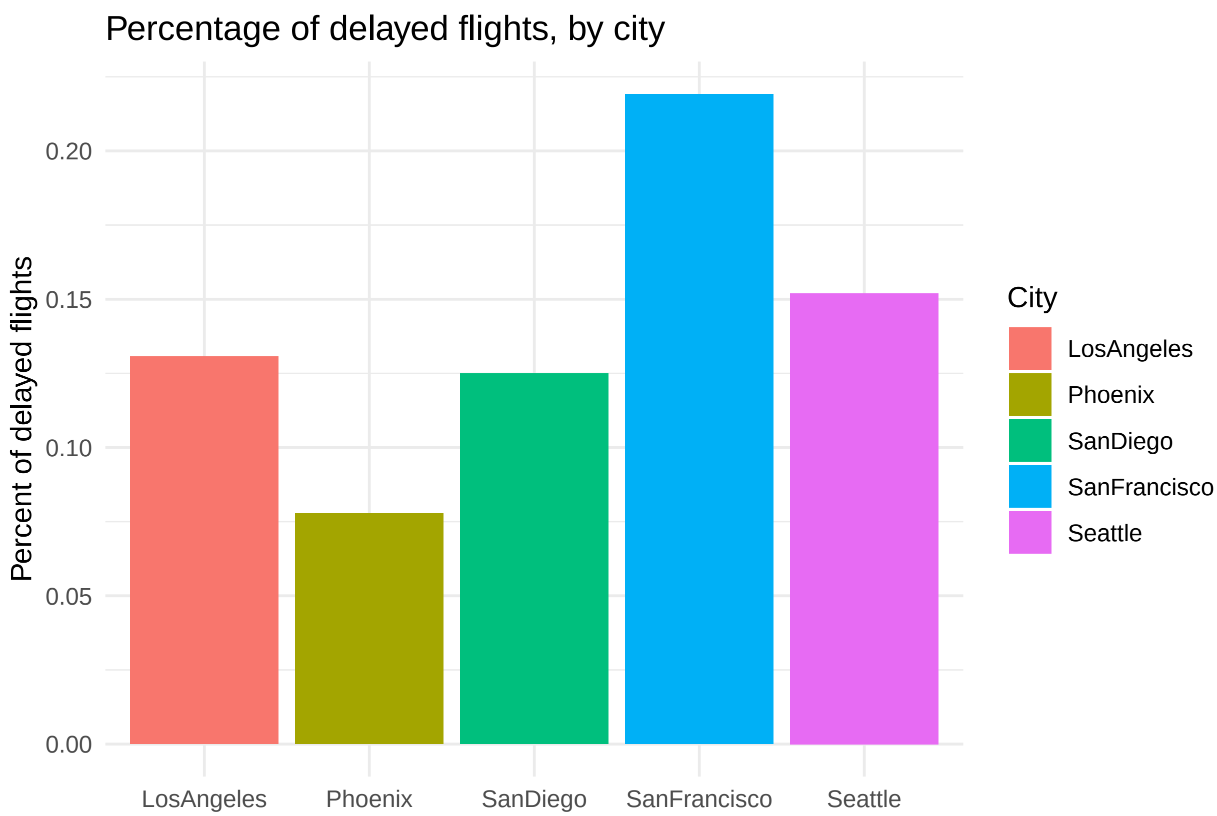
The above helps clarify the picture.

Phoenix is known to be a city with "good weather", where it seldom rains.

AM WEST (which no longer exists as an independent entity due to its 2005 merger with US Air, which then merged in 2015 with American Airline) was based in Phoenix, which is where it dominated the market. It also had larger market share (than ALASKA) in both San Diego and Los Angeles, both comparatively "good weather" cities.

On the other hand, ALASKA Airlines is based in Seattle, *which is cloudy/rainy for more than 300 days per year* (the exception being July and August.) Alaska Airlines flies mainly up and down the west coast, including flights to Alaska (hence its name) plus San Francisco, where it had larger market share than AM West. *San Francisco is known for being foggy much of the time*, which results in a large percentage of flight delays there.

```
select(.data = ResultsByCity, City=City, PctDelayed=PctDelayed) %>%
    ggplot(data = ., aes(factor(City), PctDelayed, fill = City)) +
    geom_bar(stat="identity", position = "dodge") +
    theme_minimal()+
    labs( x="", y="Percent of delayed flights") +
    ggtitle("Percentage of delayed flights, by city")
```



Conclusion:

The explanation for the paradox, "*How could one airline* (ALASKA) *have better on-time performance at each city*, *while the other airline* (AM WEST) *has better on-time performance overall?*" is found in the nature of cities in which each airline chooses to predominate, and the respective propensity for delays in such cities.

An airline which flies mainly to "bad weather" locations like Seattle and San Francisco, where a larger percentage of flights experience delays, is likely to have worse *overall* on-time performance when compared against an airline which flies mostly to "good weather" cities like Phoenix, Los Angeles, and San Diego.

Even if an airline boasts better on-time performance at each city, its overall performance can suffer because of its route map.

In this regard, ALASKA has won each of the "*battles*" (based on *within-city* comparison), but AM WEST has won the "*war*."