Downloaded Airline delay data from <http://stat-computing.org/dataexpo/2009/the-data.html.This> relates to airline delay in the year 2008.

classes <- c(year="integer",

month="integer",

dayofmonth="integer",

dayofweek="integer",

deptime="integer",

crsdeptime="integer",

arrtime="integer",

crsarrtime="integer",

uniquecarrier="character",

flightnum="integer",

tailnum="character",

actualelapsedtime="integer",

crselapsedtime="integer",

airtime="integer",

airdelay="integer",

depdelay="integer",

origin="character",

dest="character",

distance="integer",

taxin="integer",

taxiout="integer",

cancelled="integer",

cancellationcode="character",

diverted="integer",

carrierdelay="integer",

weatherdelay="integer",

nasdelay="integer",

securitydelay="integer",

lateaircraftdelay="integer")

1.MERGED ANOTHER TABLE WITH THE CURRENT TABLE IN ORDER TO FETCH AIRLINE NAME.

2.INSERTED THAN INFORMATION INTO THE CURRENT TABLE.

3. LINK USED TO CONNECT BOTH TABLES IS BY PICKING UP “UNIQUE CARRIER” FROM CURRENT TABLE AND “CODE” FROM SECOND TABLE

merged = merge(Github, carriers, by.x = "UniqueCarrier", by.y = "Code")

NEXT MOTIVE IS TO FIND HOW MANY CARRIERS DOES EACH AIRLINE HAVE WHICH WILL GIVE US AN IDEA OF THE AIRLINE WITH MAXIMUM CARRIERS

1.INSTALLED RSQLITE FOR THIS PURPOSE USING THE BELOW COMMANDS

f = dbGetQuery( conn , " SELECT DESCRIPTION, COUNT(UNIQUECARRIER) , UNIQUECARRIER FROM MERGED GROUP BY DESCRIPTION")

install.packages("RSQLite")

# CREATED A FILE NAME MERGED

conn = dbConnect(SQlite() , "merged.db")

#CREATE TABLE IN DB

dbWriteTable(conn, "merged", merged)

#LIST FIELDS IN THE TABLE

dbListFields(conn , "merged")

# WE ARE FINDING THE NUMBER OF FLIGHTS EACH AIRLINE HAS USING SQL

f = dbGetQuery( conn , " SELECT DESCRIPTION, COUNT( DISTINCT FLIGHTNUM) , FLIGHTNUM, UNIQUECARRIER FROM MERGED GROUP BY DESCRIPTION")

# WE ARE FINDING THE SAME INFORMATION USING R

Num = aggregate(merged$FlightNum~merged$Description, data = merged , FUN = function(x) length(unique(x)))

# I WANT TO CHANGE THE NAME OF THE COLUMNs

names(Num)[2]= "flight count"

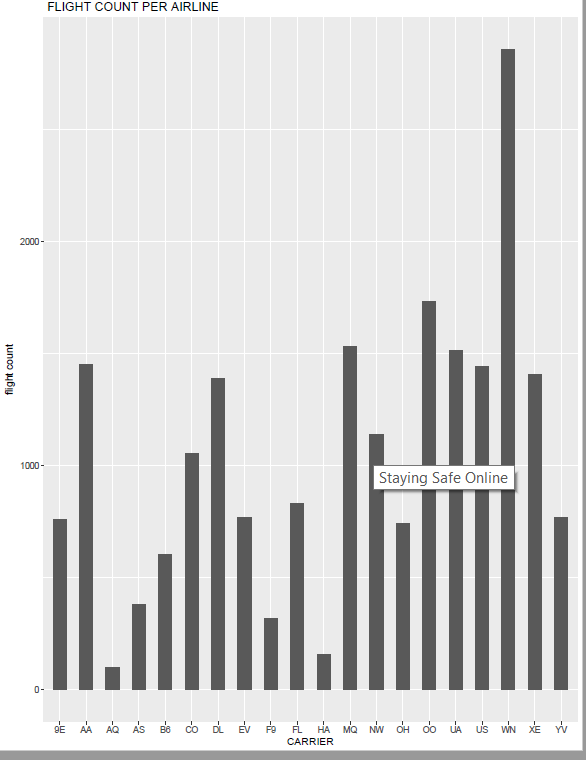
names(Num)[1] ="Description"

# ADDING COLUMN IN OBJECT NUM BY PICKING UP UNIQUE CARRIER VALUES FROM OBJECT f IN ORDER TO GET UNIQUE CARRIERS OF THE FLIGHT NUMBERS

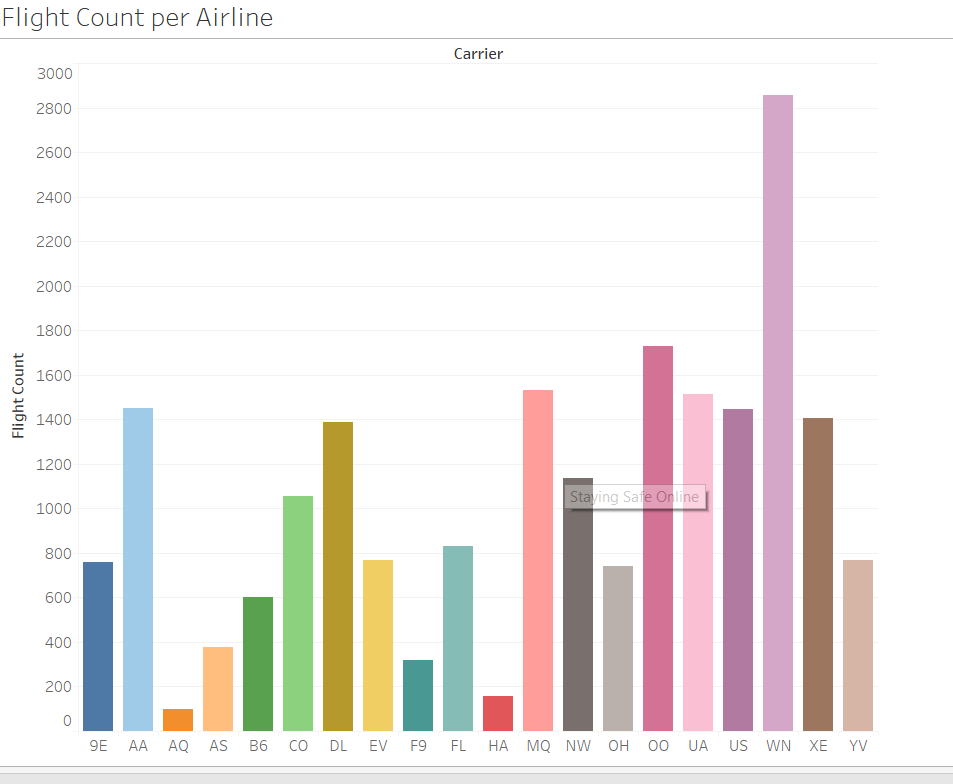
Num$CARRIER<-f$UniqueCarrier

# VISUALIZING THIS IN R USING GGPLOT

p = ggplot(Num, aes(x = CARRIER, y = `flight count`))+geom\_bar(stat= "identity", width = 0.5, position = "stack")+ggtitle (" FLIGHT COUNT PER AIRLINE")



# VISUALZING SAME DATA IN TABLEAU



# CALCULATE TOTAL DISTANCE TRAVELLED BY EACH AIRLINE THROUGHOUT THE YEAR

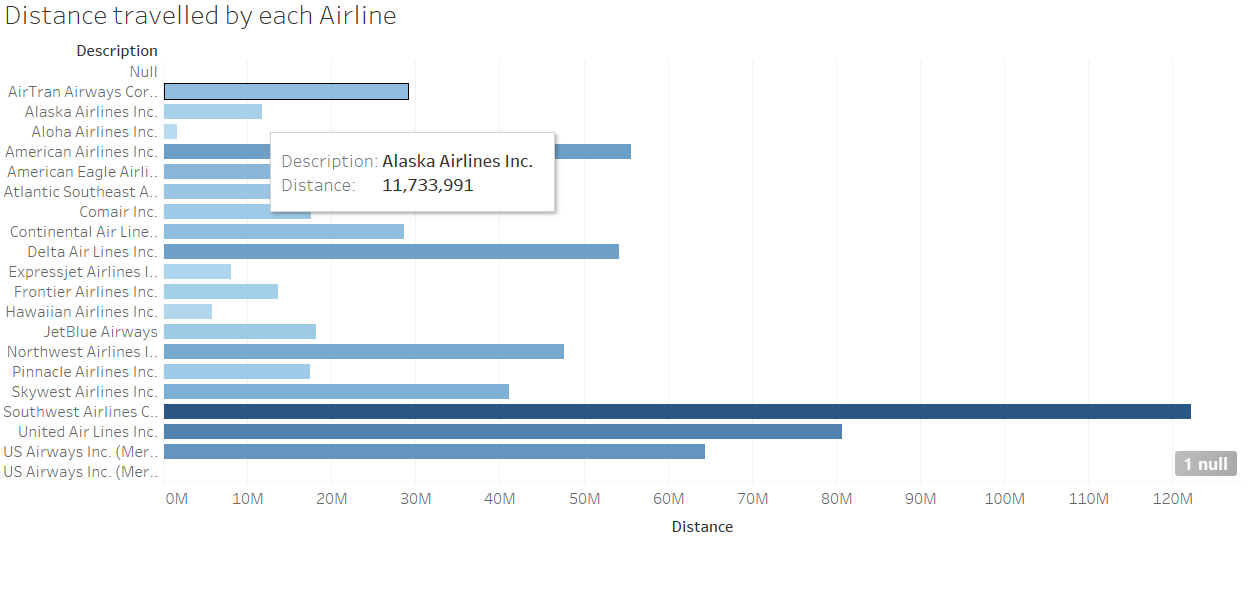
c = aggregate(Distance ~ Description , merged , sum)

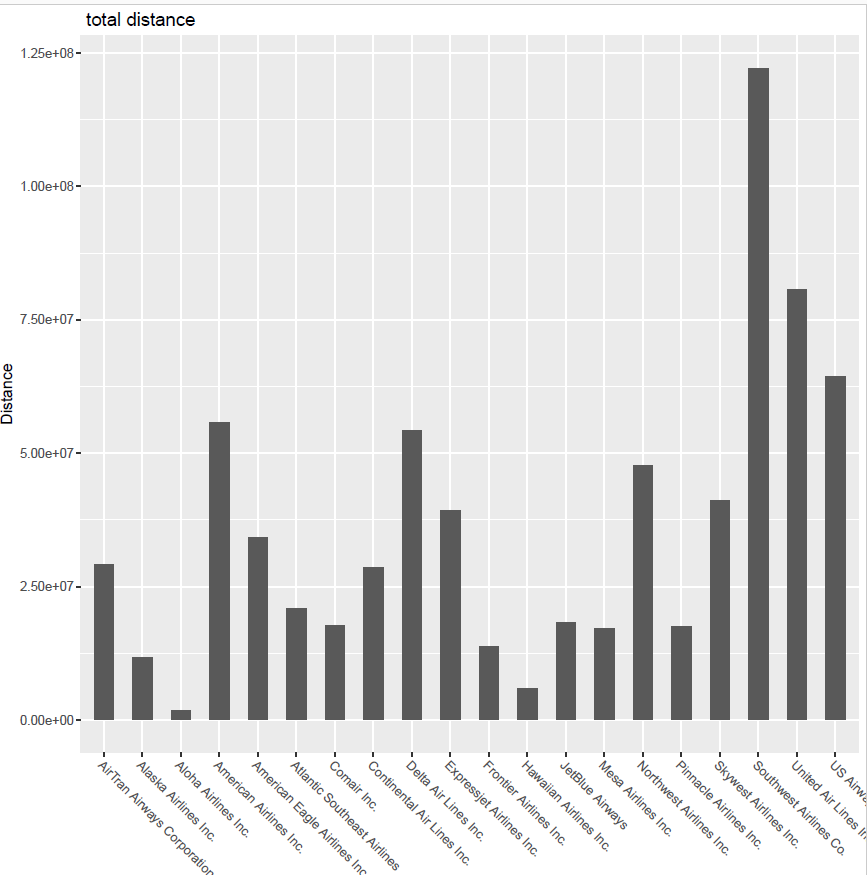
OR

X = dbGetQuery( conn , "SELECT SUM(DISTANCE), DESCRIPTION FROM MERGED GROUP BY DESCRIPTION")

ggplot(c , aes( x = Description, y = Distance))+geom\_bar(stat ="identity", width =0.5, position ="stack")+ggtitle(" total distance")+theme(axis.text.x =element\_text(angle = -45, hjust = 0, vjust= 1))

ggplot(c , aes( x = Description, y = Distance))+geom\_bar(stat ="identity", width =0.5, position ="stack")+y\_continuous(limits = c(1000000, 10000000) , breaks=c( 1000000, 10000000, 1000000))+ggtitle(" total distance")+theme(axis.text.x =element\_text(angle = -45, hjust = 0, vjust= 1))



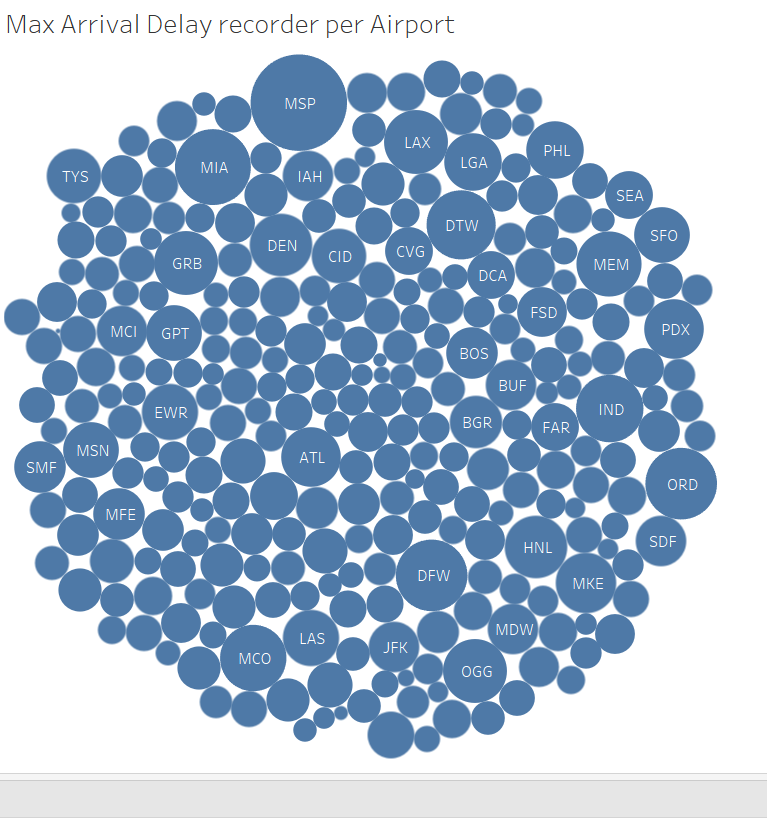


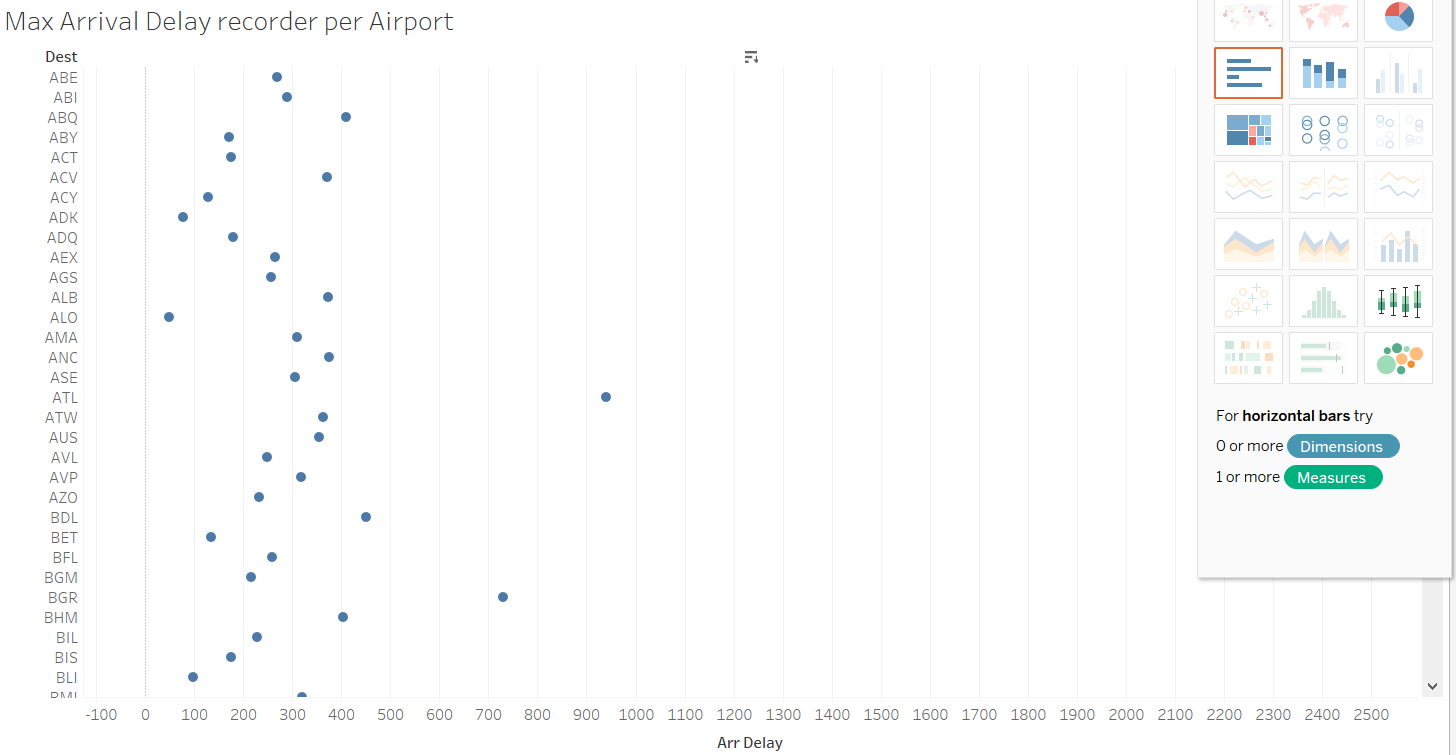
# HOW TO FIND DELAY DETAILS max arrival delay per airport

e = aggregate(ArrDelay~Description, merged, function(x) max(x))

|  |
| --- |
|  |
| |  | | --- | | > | |

ggplot(e , aes( x = Description, y = ArrDelay))+geom\_bar(stat ="identity", width =0.5, position ="stack")+ggtitle(" Arrival Delay")+theme(axis.text.x =element\_text(angle = -45, hjust = 0, vjust= 1))

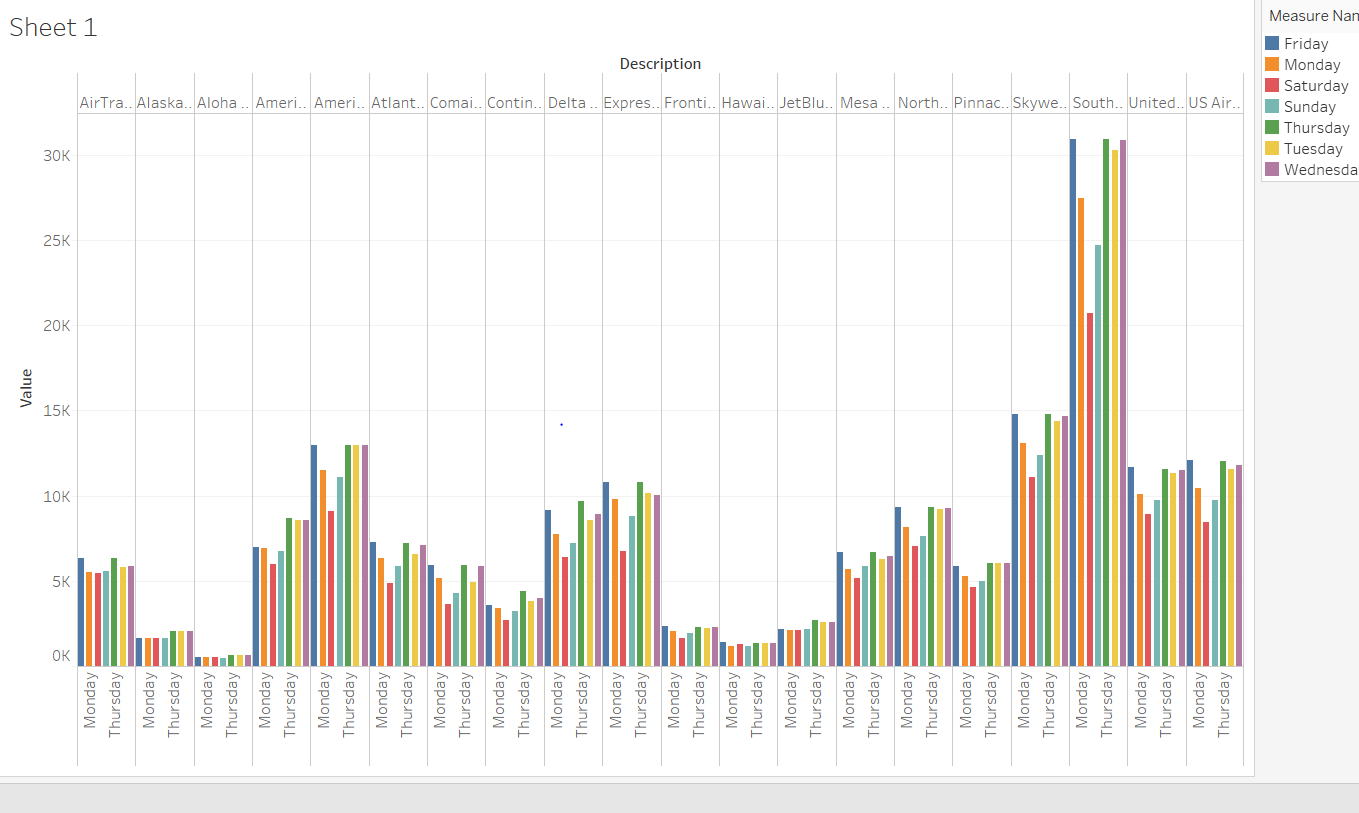


# Visualised with Tableau 

# Analyse which airline has maximum number of take offs from any particular Origin City.

jij = table(Description, DayOfWeek, Origin )

> g= xtabs(~Description+ DayOfWeek+Origin , data = jij)



#How to find the origin City that is most frequently used for travel

We see that all Airports are equally visited by every airline during seven days of the week

h = xtabs(~Origin+Description , data = jij)

mew = merge( h, iata , by.x= "iata", by.y = "Origin")

