

TEAM BIKE - ANALYZING CITI BIKES IN NYC AND PREDICTING BIKE USAGE

By Julian DeGroot-Lutzner and Vikramaditya Salwan

BACKGROUND & MOTIVATION

WHAT IS CITI BIKES?

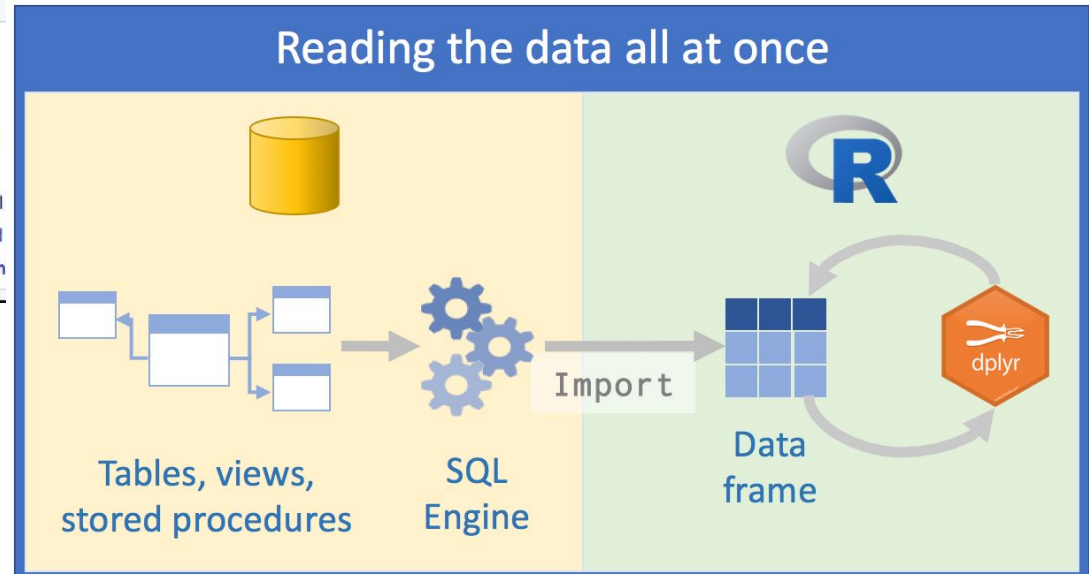
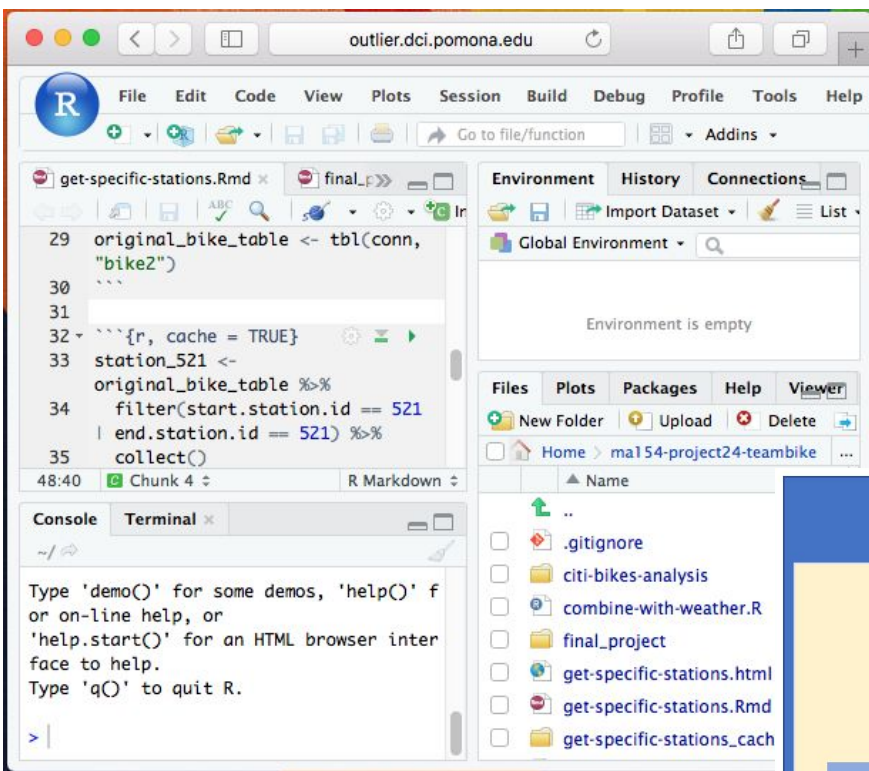
SOME OF OUR DRIVING QUESTIONS

- DOES WEATHER AFFECT BIKE USAGE?
- CITI BIKES' TARGET MARKET?
- WHAT TIME ARE MORE BIKES GETTING USED?
- WHAT STATIONS ARE GETTING THE MOST TRAFFIC?
- ARE THERE INCONSISTENCIES IN STATIONS - GETTING MORE INFLOW THAN OUTFLOW OR VICE VERSA?

19 Jan

New Visitor Returning Visitor





Lifecycle of an Analysis Project

Subset

Extract data to explore, work with

Clarify

Become familiar with the data,
template a solution

Develop

Create a working model

Productize

Automate and integrate

Publish

Socialize

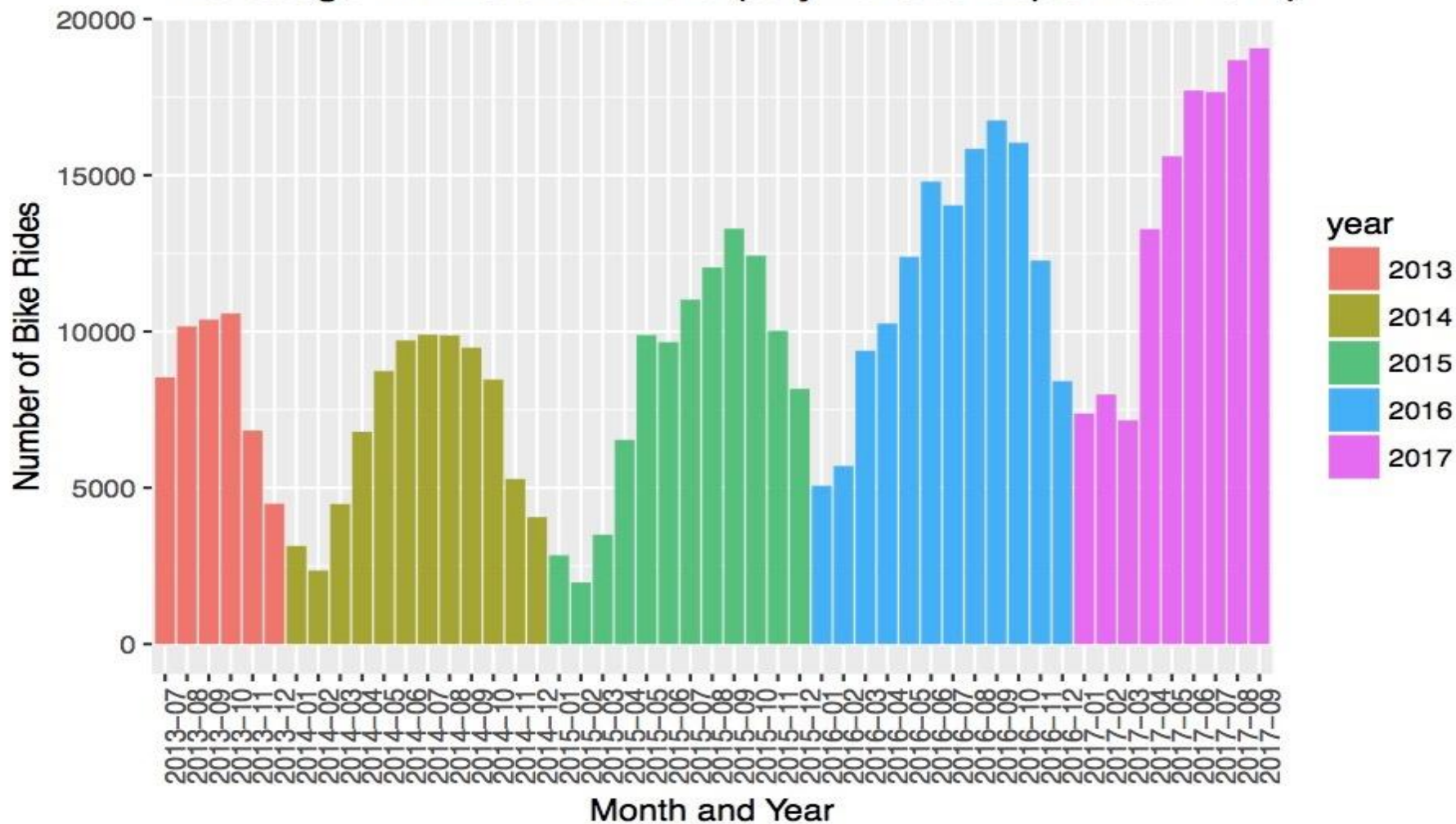
Class 1



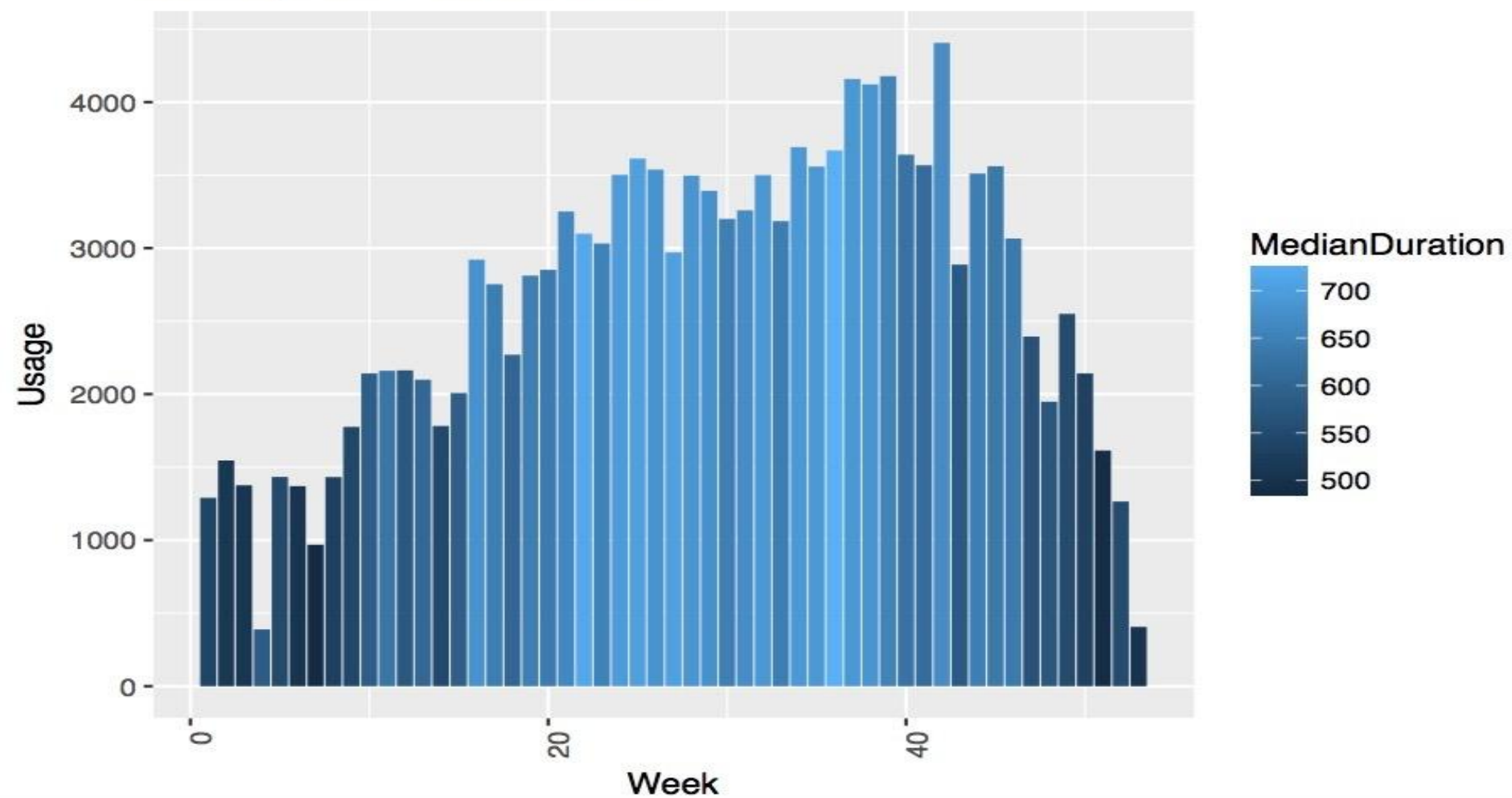
Class 2
Class 3



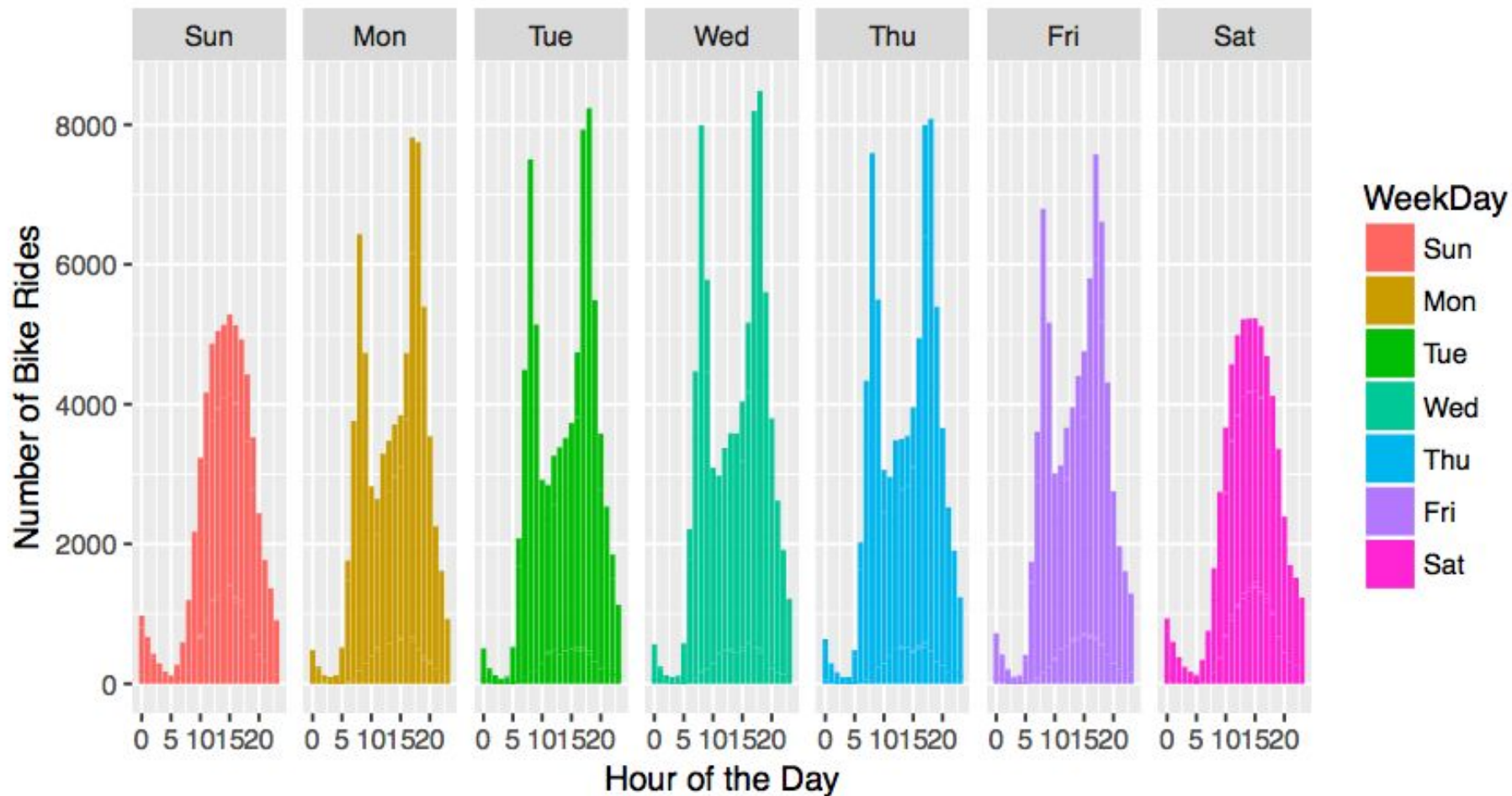
Bike Usage in Different Months (July 2013 to September 2017)



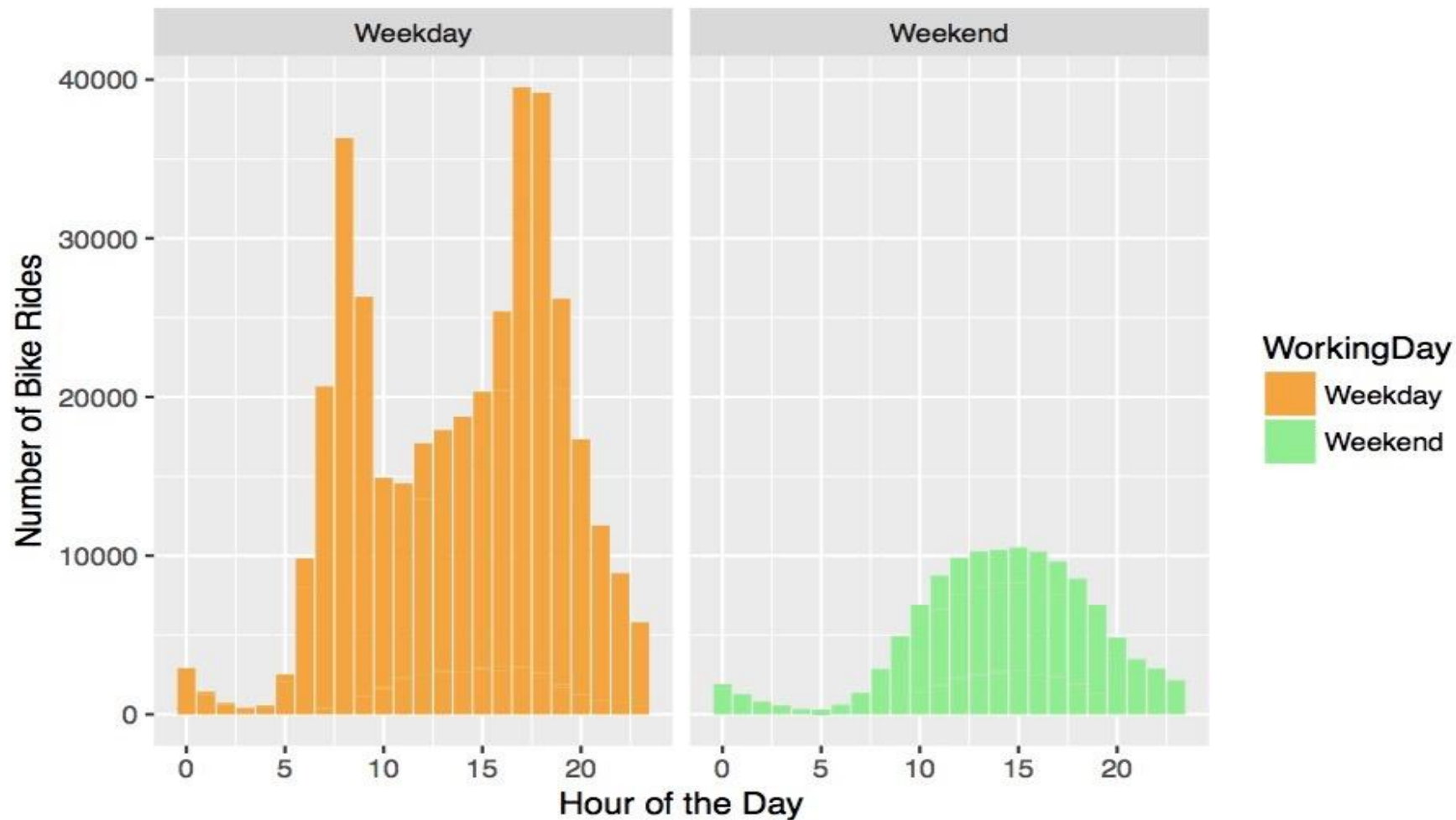
Weekly Bike Rides in 2016



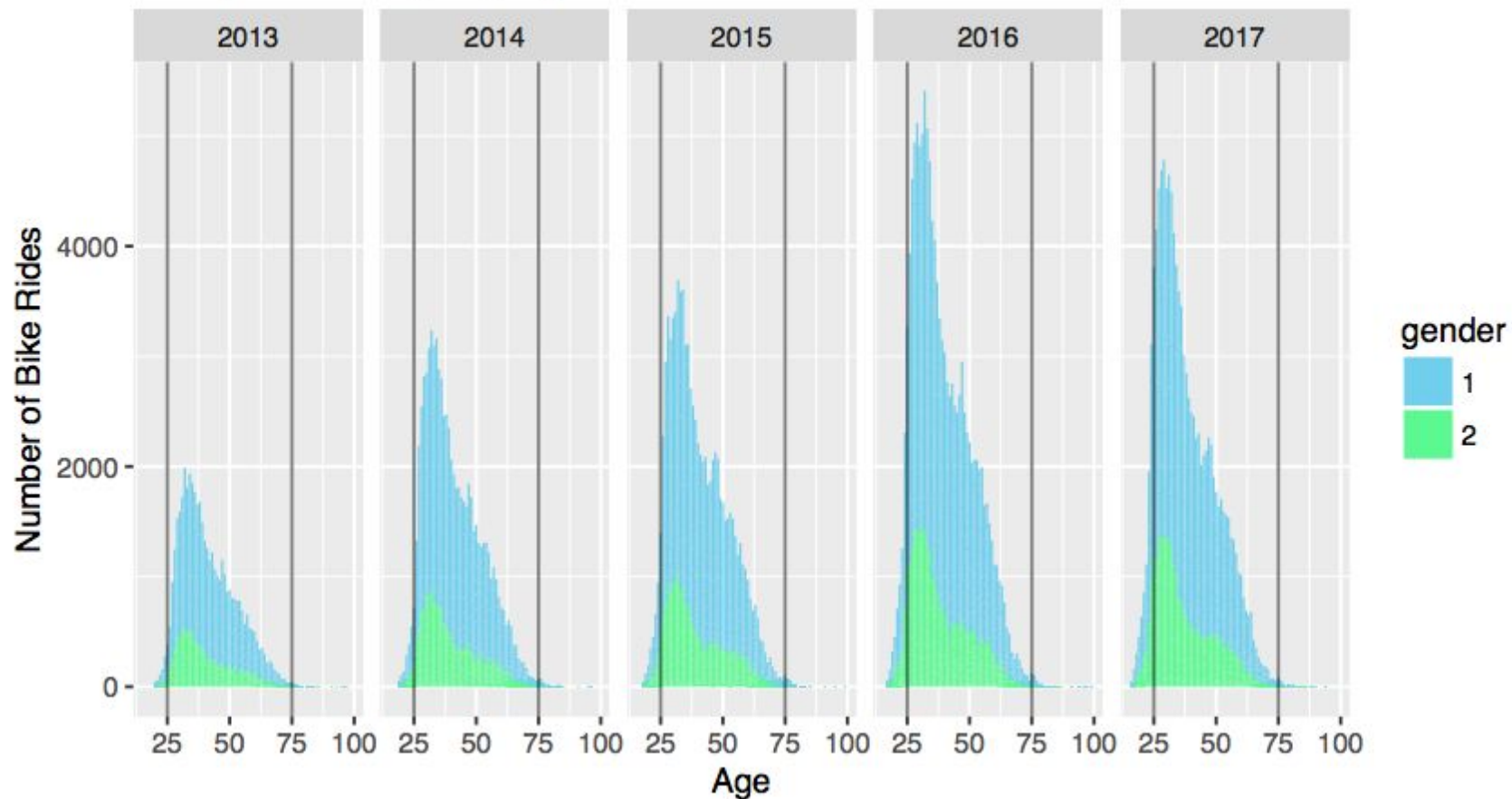
Bike Usage on Different Days



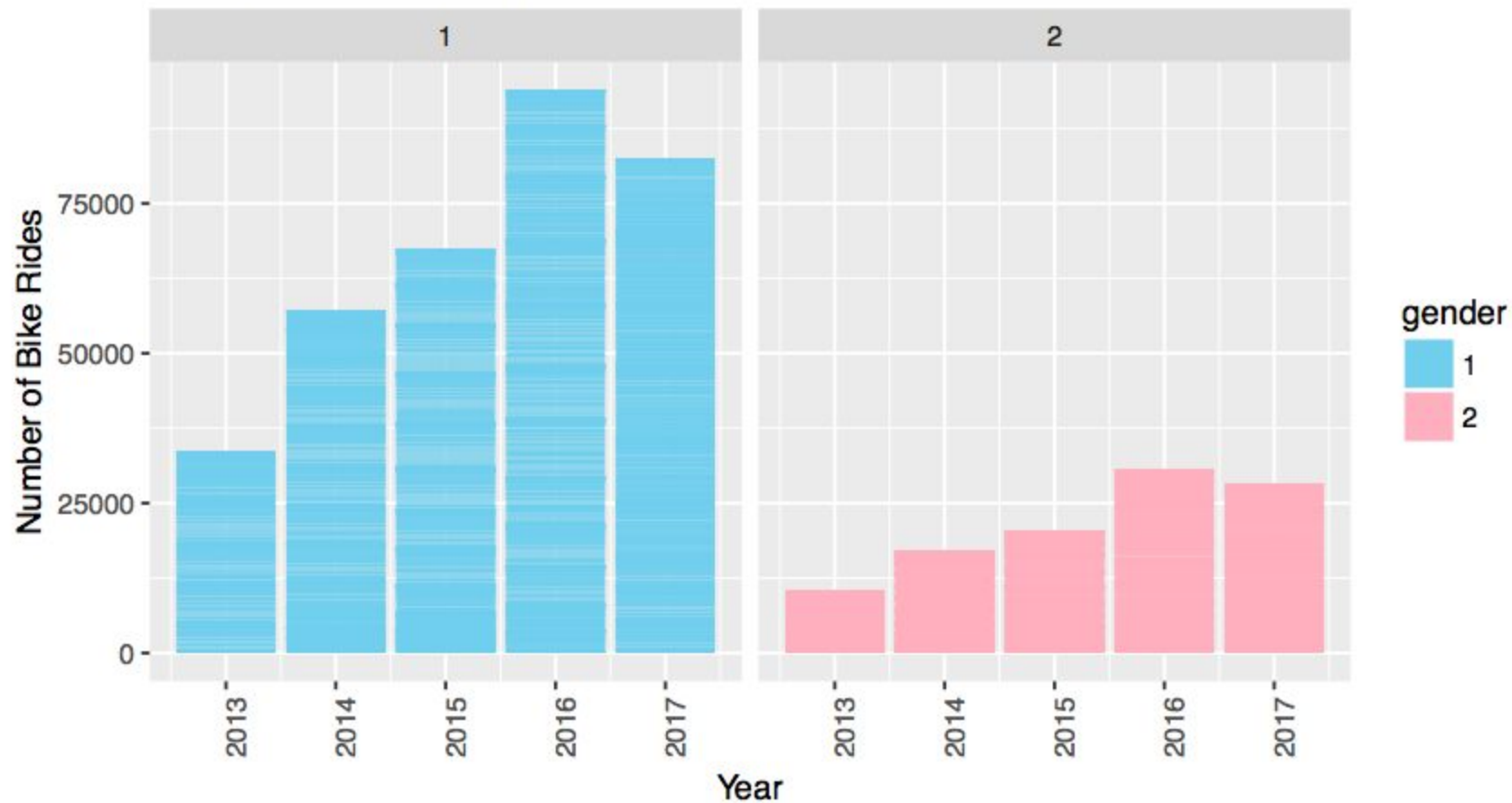
Bike Usage on Weekdays versus Weekends



Yearly Bike Usage by Age and Gender



Yearly Bike Usage by Gender



EXPLORATORY ANALYSIS - KEY TAKEAWAYS

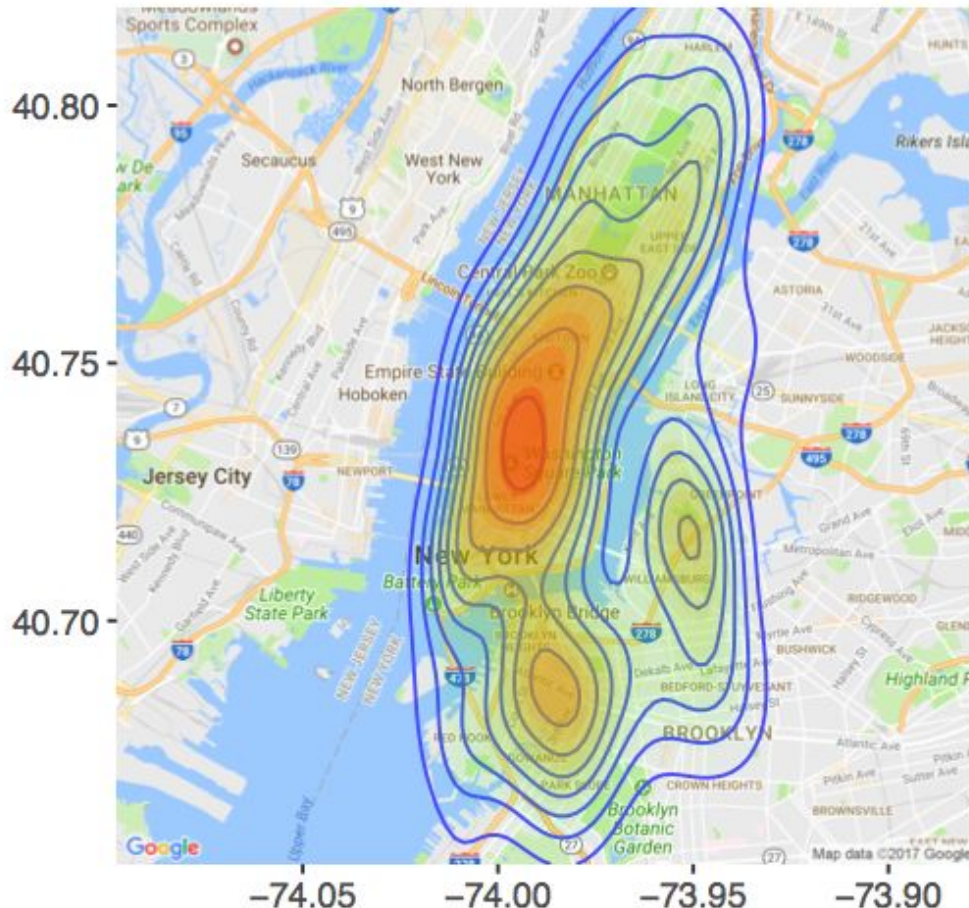
BIKE SHARING HAS BEEN A MAJOR SUCCESS.

WEATHER CAN PLAY A SPOILSPORT.

FEMALES NOT USING BIKES RELATIVE TO MALES.

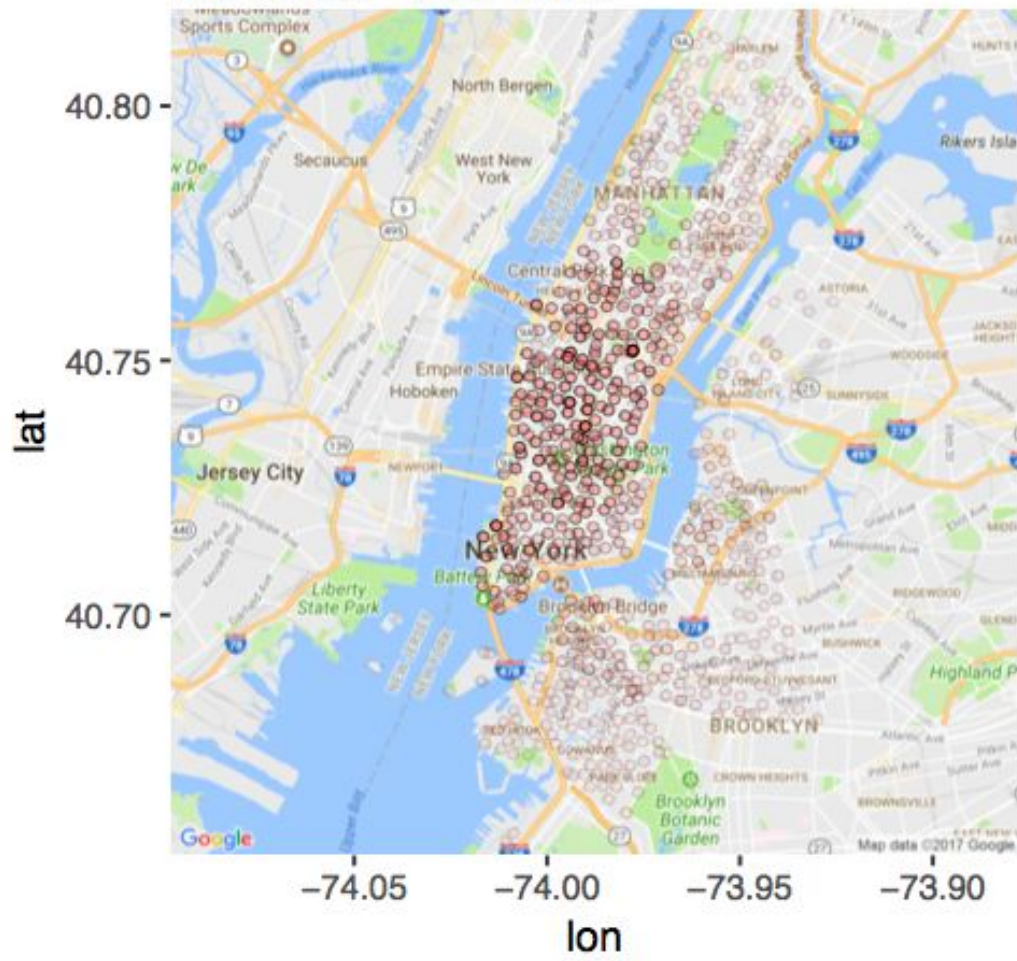
TARGET MARKET - COMMUTERS AND 25 TO 50 YEAR OLDS.

The map displays population density contours for New York City. The highest density is concentrated in the center of Manhattan, around Central Park and the Empire State Building, indicated by red and orange colors. The density decreases as one moves away from the center, with green and blue colors representing lower density areas. The map includes labels for various neighborhoods and landmarks, such as Central Park, Empire State Building, and the Brooklyn Bridge. The map is overlaid with a grid of latitude and longitude coordinates.



Number of Rides per Station

Darker Red = More Rides



IMPORTERS AND EXPORTERS

TOP NORMALIZED IMPORTERS

```
## # A tibble: 10 x 5
##       id start.count end.count difference normalized.difference
##   <int>      <int>    <int>      <int>          <dbl>
## 1    521      2649     2380        269          0.05349
## 2    519      5367     4835        532          0.05215
## 3    281      2493     2279        214          0.04484
## 4    490      3264     3064        200          0.03161
## 5    479      2026     1904        122          0.03104
## 6    457      2287     2160        127          0.02856
## 7    523      2631     2496        135          0.02633
## 8   2006      2706     2574        132          0.02500
## 9    380      2199     2105         94          0.02184
## 10   528      1889     1811         78          0.02108
```

STATION 521

-2649

+2380

STATION
521

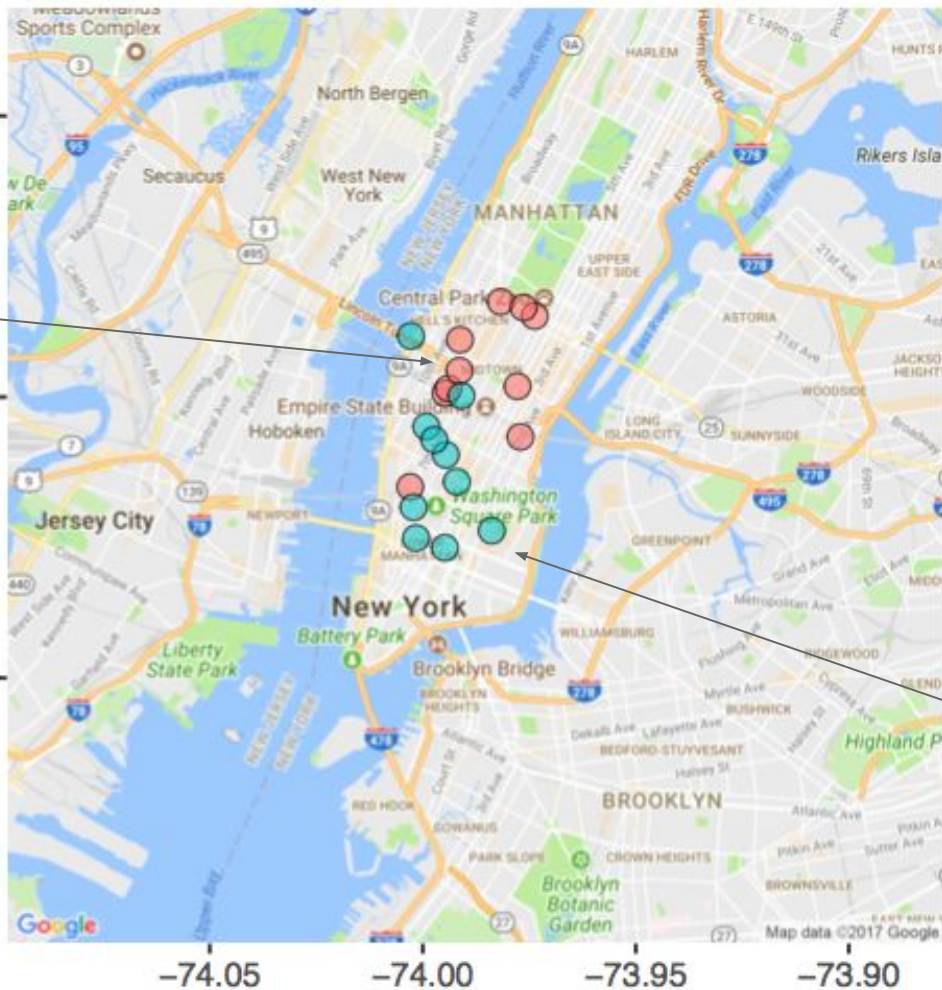
EXPORTER

lat

40.80 -

40.75 -

40.70 -



STATION
432

IMPORTER

lon

-74.05

-74.00

-73.95

-73.90

RANDOM FOREST

OUR MODEL

Confusion Matrix and Statistics

##

Reference

Prediction FALSE TRUE

FALSE 2232 918

TRUE 1239 1941

##

Accuracy : 0.659

BASELINE

Confusion Matrix and Statistics

##

Reference

Prediction FALSE TRUE

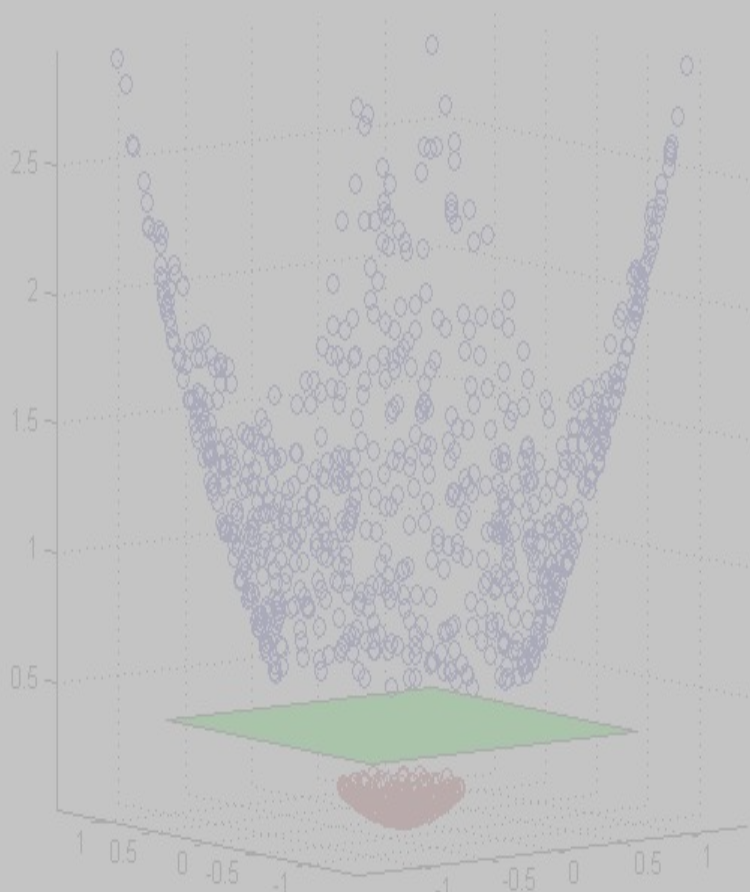
FALSE 3471 2859

TRUE 0 0

##

Accuracy : 0.548

RMSE = 16.86



```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    N    P
##           N 2614 1451
##           P  858 1408
##
##           Accuracy : 0.635
##           95% CI : (0.623, 0.647)
##           No Information Rate : 0.548
##           P-Value [Acc > NIR] : <2e-16
##
##           Kappa : 0.25
##           McNemar's Test P-Value : <2e-16
##
##           Sensitivity : 0.753
##
##           Pos Pred Value : 0.643
##           Neg Pred Value : 0.621
##           Prevalence : 0.548
##           Detection Rate : 0.413
##           Detection Prevalence : 0.642
##           Balanced Accuracy : 0.623
##
##           'Positive' Class : N
##
```

SUPPORT VECTOR MACHINES MODEL

A close-up photograph of a person's hands holding a purple marker, drawing on a whiteboard. The background is blurred, showing some office equipment. The text 'TAKE AWAYS' is overlaid in a white, handwritten-style font.

TAKE AWAYS

Learned to adapt:

Start with a sample

Understand the sample

Create a model

Refine the question and
repeat first steps

THANK YOU!

QUESTIONS?

Weather
Random Sample
- join w/ weather data
- visualize

