Predicting Citi Bike Usage in NYC

David Skålid Amundsen

March 15, 2016

Goal: Predict Citi Bike Usage in NYC

- Demand varies with time, but what determines it?
- What is demand?
  - Average trip length.
  - Number of bikes on the road.
  - Number of new bike rentals per time.
- I will try to predict:
  - Number of new bike rentals per time.
  - Predictors: **time** and **weather**.

Citi Bike Data

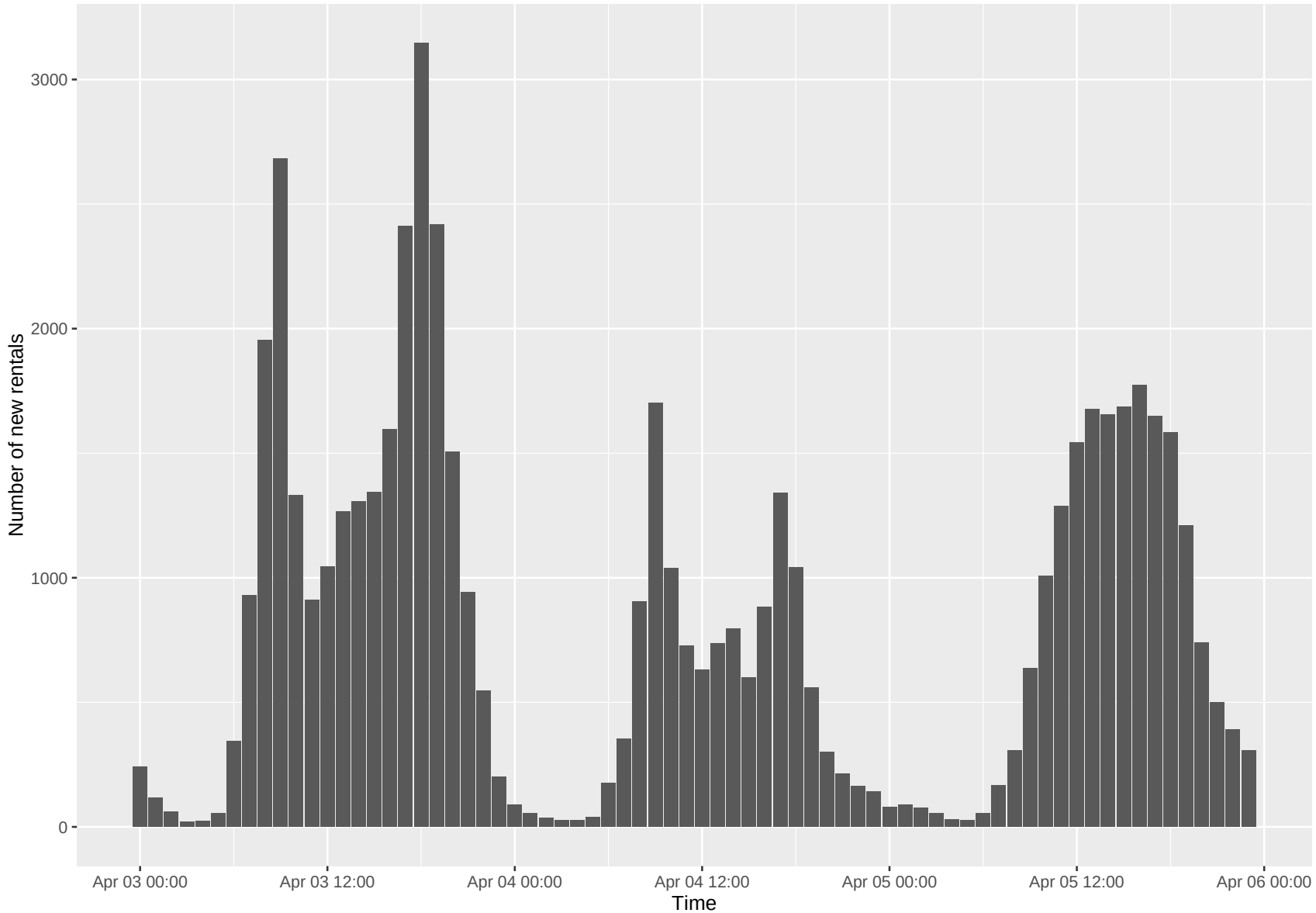
Details each trip:

- Duration
- Start and end time
- Stations
- Age
- Gender
- Bike ID.

I reduce it to:

- Time
- Number of new bike rentals

Citi Bike Usage on July 4th 2015

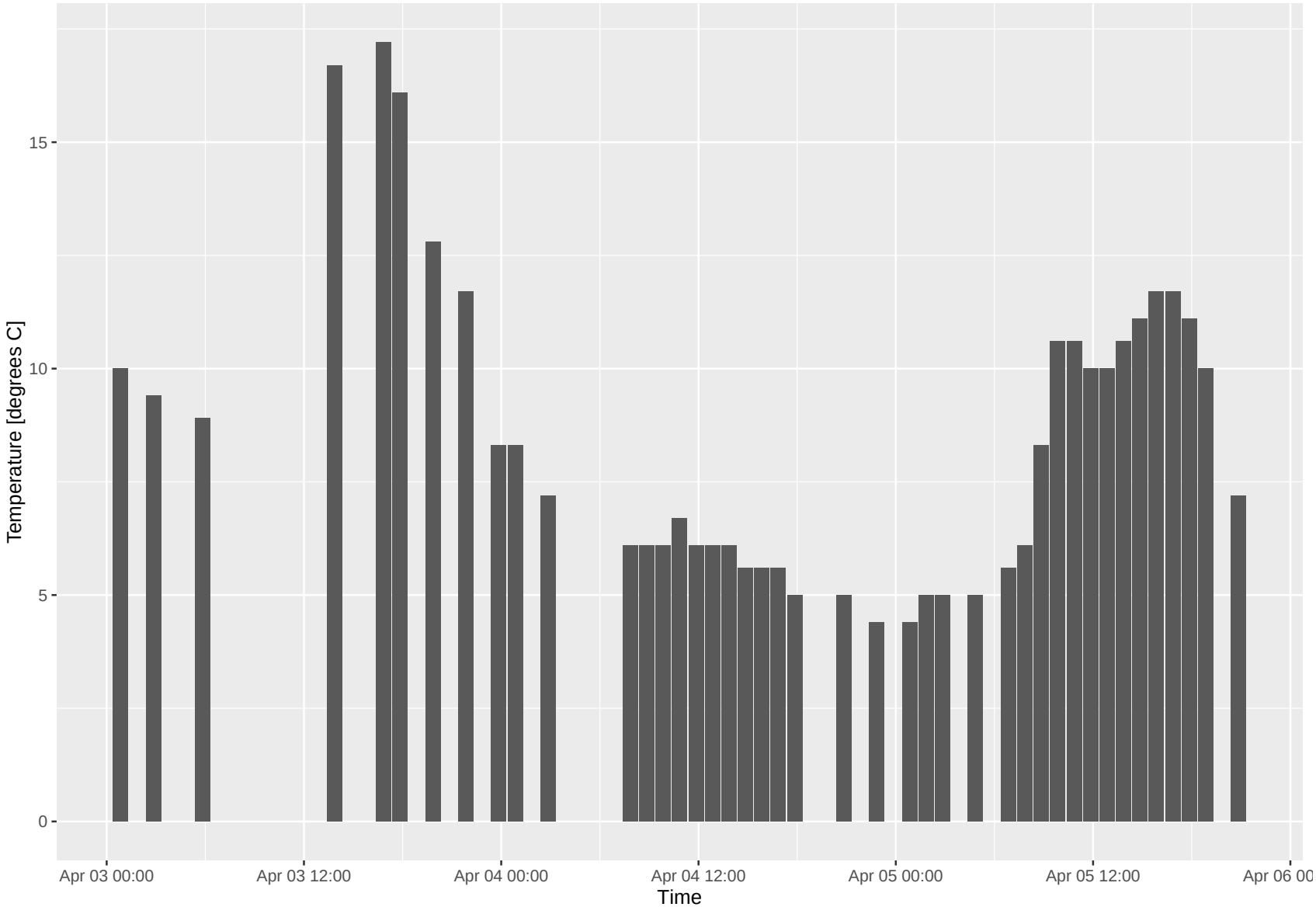


Weather Data

Weather observations from NOAA's weather station in central park:

- Wind speed
- Direction
- Temperature
- Precipitation
- Pressure
- Dew point

Temperature observations



Data for modelling

Using the following predictors:

- Weather:
  - Wind speed and direction
  - Temperature
  - Precipitation
  - Pressure
  - Dew point.
- Time:
  - Hour of day
  - Weekday

Consequence:

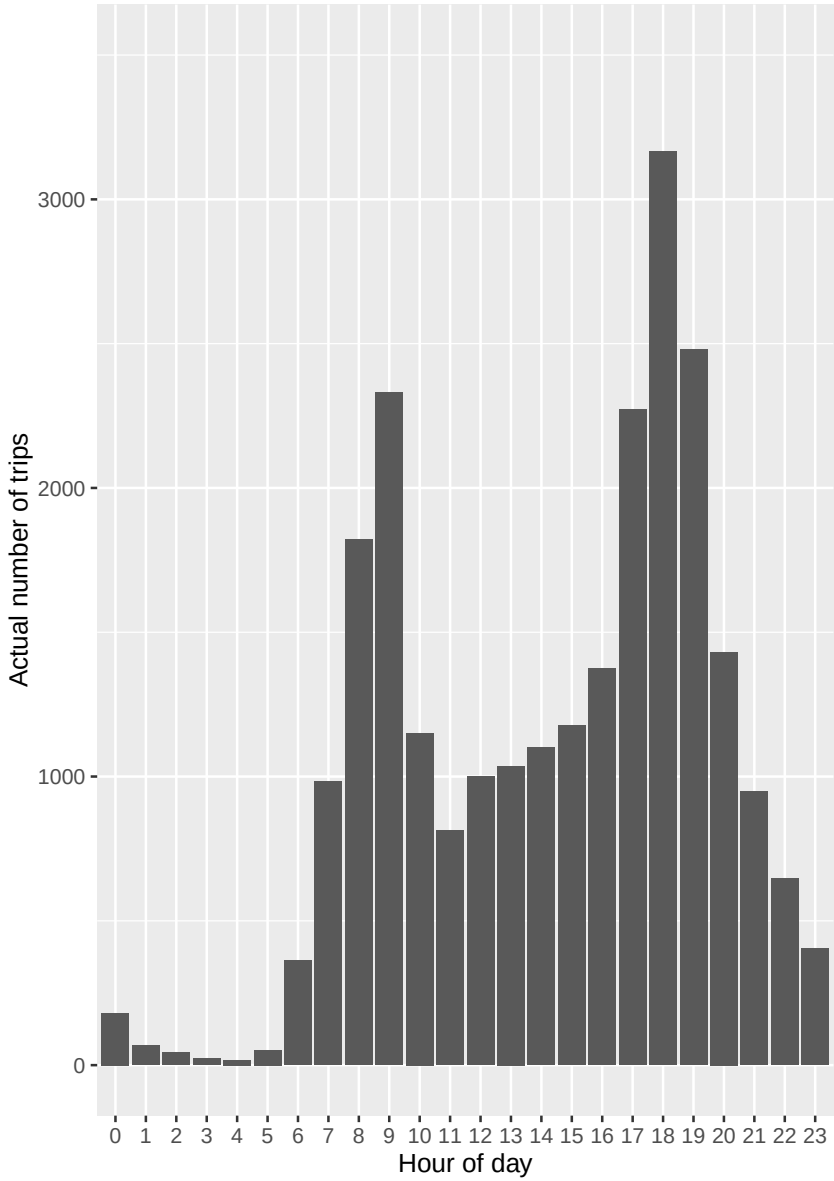
- Any seasonal variations are modelled by weather.

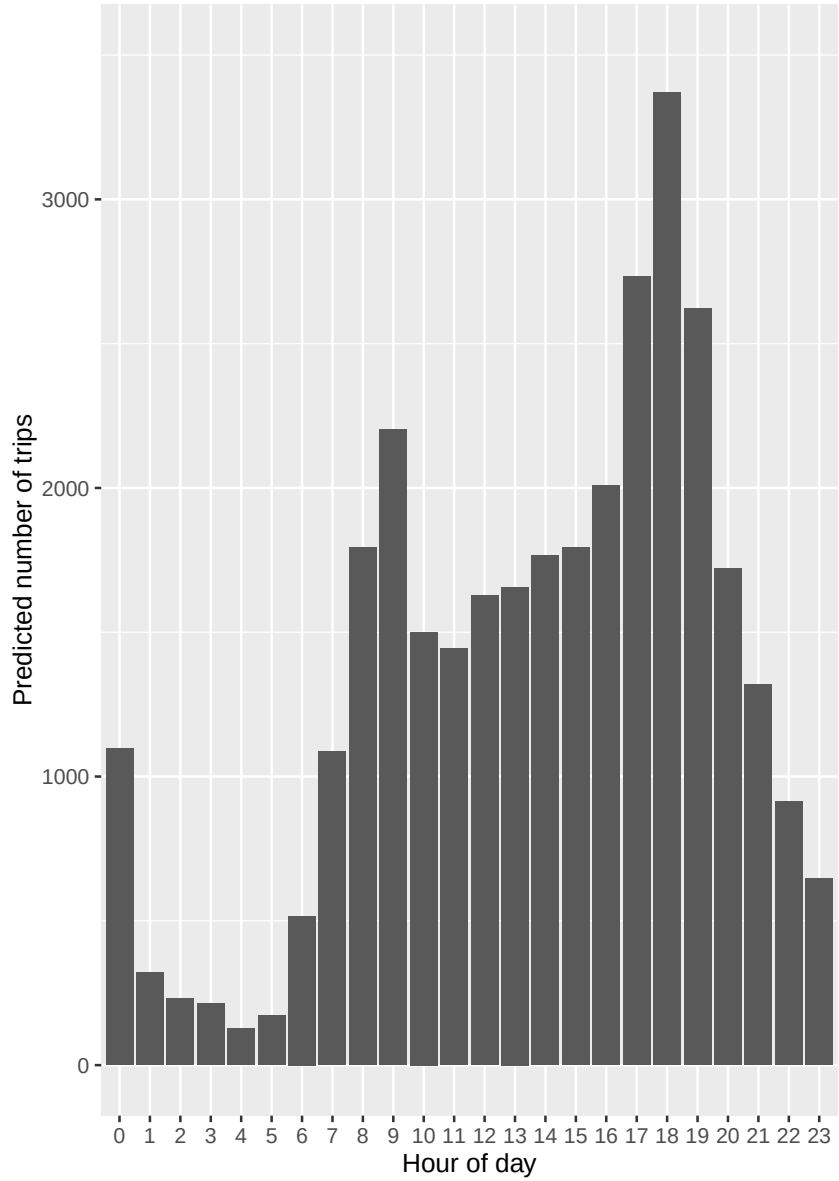
Model result:

Random forest provided best fit and prediction:

- RMSE: 360 trips per hour
- R^2: 0.85
- RMSE on test set: 362 trips per hour

Test prediction on random day in April 2015 (13th)





Further improvments

- Depends on major events?
- Yearly trends?
- Depends on station location?
- Depends on cloud cover?

Other models applied

Generalized linear model:

- RMSE of fit: 497 trips per hour
- R^2: 0.72
- RMSE of test set: 506 trips per hour

Regression tree:

- RMSE of fit: 419 trips per hour
- R^2: 0.80
- RMSE of test set: 414 trips per hour