01-introduction

02-data

Read data from website

Unzip to csv

Load csv to database mysql ??

03-cleaning

Basic data analysis (outliers) --Gaoyi

Check 3 month (201806,201906,202006)

04-missing

Pattern before removing outliers

Pattern after removing outliers

05-results -by Monday night (12/7)

Consistent color for histogram, barchart, boxplot…, color=’black”, fill=’lightblue’,

Data exploration

1. Age/gender vs bike usage (count, tripduration, path, speed, distance)
2. Weekday/weekend vs bike usage (count, busiest/popular paths (from A to B) )
3. travel between Manhattan and the outer boroughs, average hourly (choose one week, only weekdays)
4. weather vs bike usage (temperature, rain/snow)
5. subscriber/customer vs bike usage (simple analysis is enough)

06-interactive

07-conclusion