Analyzing Freight Congestion and Transportation Performance Measures Using National Performance Measurement Research Data Set (NPMRDS)

Author: Yaxin Zhang
Supervisor: Yunhai Zhang(Original)
Jan Khan (NYMTC)
Ali Afshar (NYMTC)
Prof. Alison J. Conway (CCNY)

Presentation Outline

- Part I. Introduction to NPMRDS
- Part II. Methodology and Performance Measurements
- Part III. Clustering and Machine Learning
- Part IV. Dashboard Display

Part I. National Performance Measurement Research Data Set (NPMRDS)

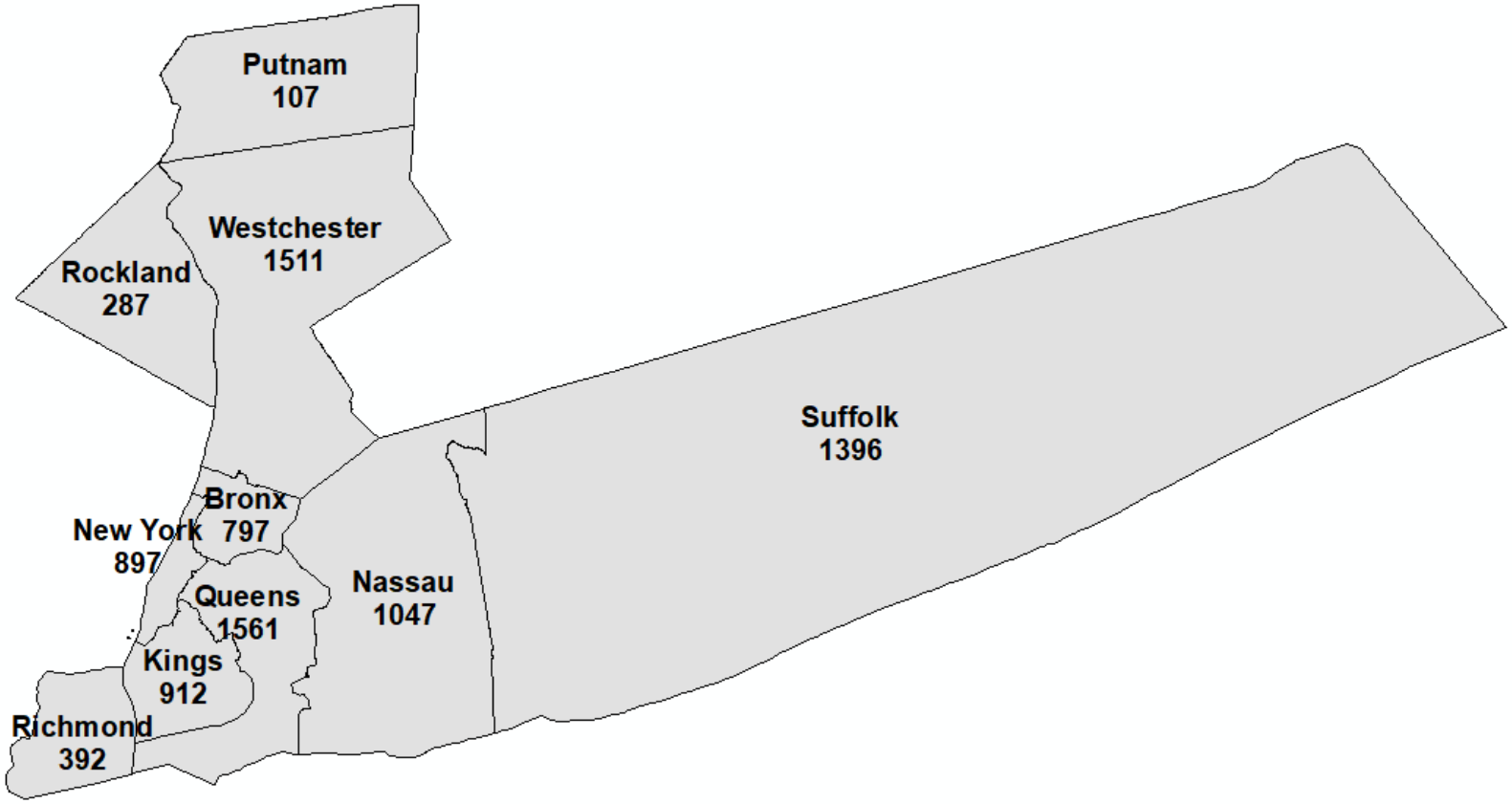
- Procured and sponsored by the **Federal Highway Administration (FHWA)**
- Records a national dataset of **average travel time every 5 minutes based on road segments called Traffic Message Channels (TMC)**
- Consists of three data sets: **Passenger vehicles, Freight vehicles, Freight and Passenger vehicles combined.**
- Passenger probe data: collected from mobile phones, vehicle navigation systems and portable navigation devices.
- Freight probe data: obtained from the American Transportation Research Institute (ATRI) in fleet data collection.
- Combined freight and passenger vehicle data comes from a weighted average of passenger and freight vehicles based on respective traffic volumes.

NPMRDS

In this research, both combined and freight NPMRDS in 2018 were downloaded based on **1-hour** and **15-min epoch** covering 10 counties of New York----**New York (Manhattan), Kings (Brooklyn), Queens, Bronx, Richmond (Staten Island), Putnam, Westchester, Suffolk and Nassau**.

tmc_code	county	miles	aadt	measurement_tstamp	speed	reference_speed	travel_time_seconds
<chr>	<chr>	<dbl>	<int>	<chr>	<dbl>	<int>	<dbl>
120N17167	NASSAU	0.0339	32120	2018-01-01 02:00:00	67	52	1.82
120N17167	NASSAU	0.0339	32120	2018-01-01 04:00:00	34	52	3.59
120N17167	NASSAU	0.0339	32120	2018-01-01 08:00:00	27	52	4.52
120N17167	NASSAU	0.0339	32120	2018-01-01 09:00:00	46	52	2.65
120N17167	NASSAU	0.0339	32120	2018-01-01 10:00:00	18.8	52	6.51
120N17167	NASSAU	0.0339	32120	2018-01-01 12:00:00	18.1	52	6.75
120N17167	NASSAU	0.0339	32120	2018-01-01 17:00:00	16	52	7.63
120N17167	NASSAU	0.0339	32120	2018-01-01 18:00:00	13	52	9.39
120N17167	NASSAU	0.0339	32120	2018-01-01 19:00:00	40	52	3.05
120N17167	NASSAU	0.0339	32120	2018-01-01 20:00:00	37	52	3.3

Research area and TMC number



- Each segment is identified by a **unique tmc-code** with information including county, miles, average annual daily traffic (aadt), time period, hourly travel speed, reference speed, and travel time seconds.
- In total, 8824 (freight dataset) and 8907 (combined dataset) unique tmc-code were found from the data set.

Part II. Methodology and Performance Measurements

Phase I.
EDA and Spatial Analysis of performance measurements
(TTI, PTI, PHED, LOTTR, TTTR)

Phase II.
Segment Grouping and Clustering
(Clustering methods: K-means, K-medoids, Hierarchical, DBscan, Mixture of Gaussian)

Phase III.
Predictive Modeling of Congested Segments
(Machine learning methods: logistics regression, decision trees, Bayesian classifiers, neural network, support vector machines, etc.)

Phase IV.
Dashboard

Performance Measurements

Main performance measurements calculated in the research: *TTI, PTI, LOTTR, PHED, TTTR.*

- **Travel Time Index (TTI)**

$$TTI = \frac{TT_{MeanCongestion}}{TT_{FreeFlow}}$$

Threshold:

1.10 < TTI < 1.50	moderate congestion
1.50 < TTI < 2.00	significant congestion
TTI > 2.00	severe congestion

- **Planning Time Index (PTI)**

$$PTI = \frac{TT_{95\%}}{TT_{FreeFlow}}$$

Threshold:

PTI < 2	“Reliable
2 < PTI < 3	“Moderate Unreliability”
PTI > 3	“Severe Unreliability

Performance Measurements

Level of Travel Time Reliability (LOTTR) Metrics

$$LOTTR_i = \frac{80^{th}\ PercentileTravelTime_i}{50^{th}\ PercentileTravelTime_i}$$

Where i represents the time period:

1. 6 a.m. - 10 a.m., weekdays
2. 10 a.m. - 4 p.m., weekdays
3. 4 p.m. - 8 p.m., weekdays
4. 6 a.m. - 8 p.m., weekends

A segment is reliable if all four LOTTR metrics are lower than 1.5.

Truck Travel Time Reliability (TTTR) metrics

$$TTTR_i = \frac{95^{th}\ PercentileTravelTime_i}{50^{th}\ PercentileTravelTime_i}$$

Where i represents the time period:

1. 6 a.m. - 10 a.m., weekdays
2. 10 a.m. - 4 p.m., weekdays
3. 4 p.m. - 8 p.m., weekdays
4. 8 p.m. - 6 a.m., all days
5. 6 a.m. – 8 p.m. weekends

The highest TTTR value of the five periods for each segment is selected as the “maximum TTTR”, which is used for calculating of the Truck Travel Time Reliability Index (also called Freight Reliability Measure).

Peak Hour Excessive Delay

Excessive Delay threshold travel time (EDTTT)

$$\left(= \frac{TravelTimeSegmentLength_s}{ThresholdSpeed_s} \times 3600 \right)$$

Travel Time Segment Delay (RSD)
per segment per time bin

$$\left(= TravelTime_{s,b} - EDTTT_{s,b} \right)$$

Excessive Delay
per segment per time bin

$$(= \begin{cases} \frac{RSD_{s,b}}{3600} & \text{When} \quad RSD_{s,b} \geq 0 \\ 0 & \text{When} \quad RSD_{s,b} \leq 0 \end{cases})$$

Total Excessive Delay
for total time bin per segment

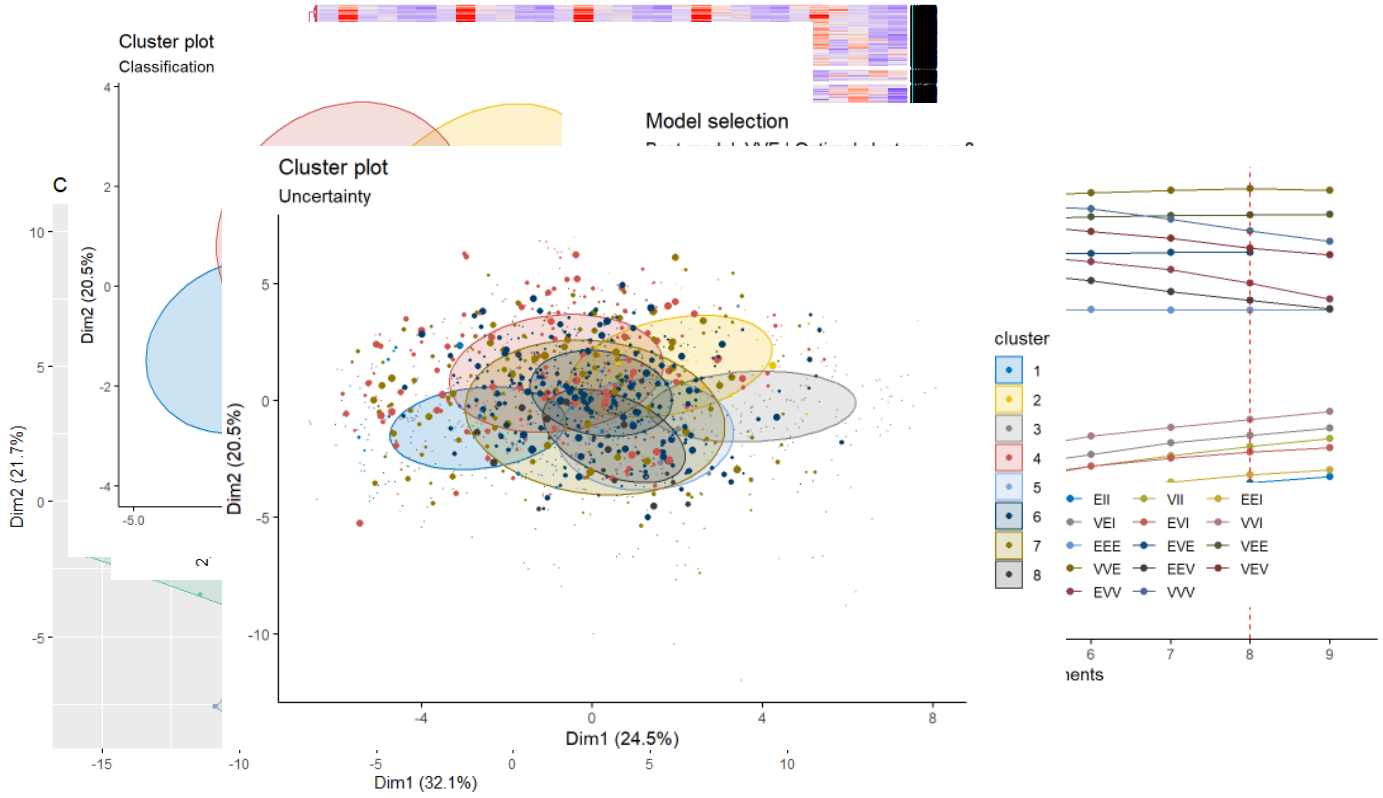
$$(= AVO \times \sum_{d=1}^{TD} \left\{ \sum_{h=1}^{TH} \left[\sum_{b=1}^{TB} \left(ExcessiveDelay_{s,b,h,d} \times \left(\frac{HourlyVolume}{4} \right)_{s,h,d} \right)_b \right]_h \right\}_d)$$

Annual Hours of Peak Hour Excessive Delay per Capita (PHED)

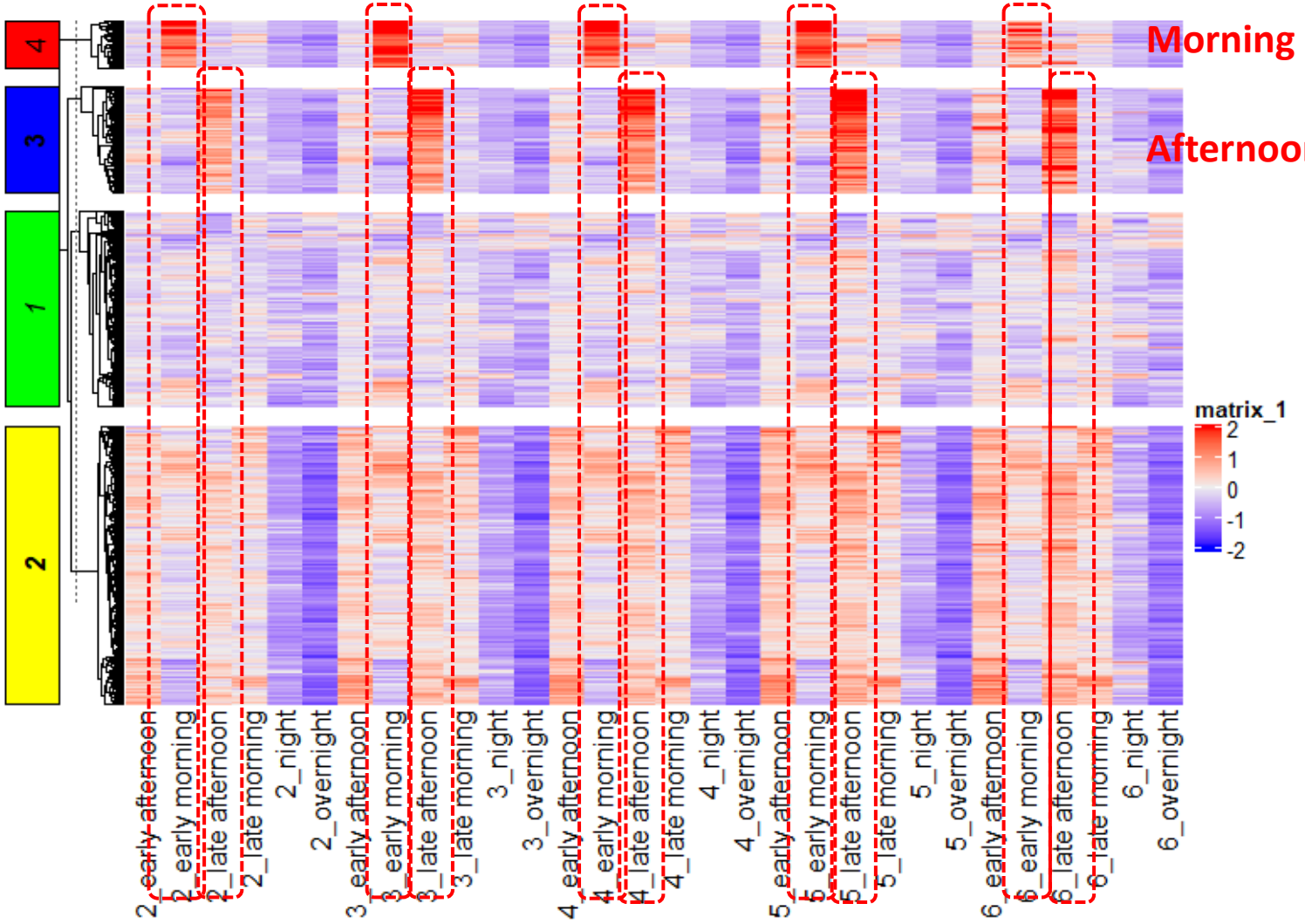
$$(= \frac{\sum_{S=1}^{T} TotalExcessiveDelay_s}{TotalPopulation})$$

Part V. Clustering

- K-means
- K-Medoids
- Hierarchical
- Dbscan
- Mixture of Gaussian



Clustering



Morning Congestion Pattern

Afternoon Congestion Pattern

Machine Learning

- ❖Linear Regression
- ❖Logistics Regression
- ❖Decision Trees
- ❖Bayesian Classifiers
- ❖Neural Network
- ❖Support Vector Machines

```
Call:
lm(formula = pm_peak ~ aadt + crash + res_d + office_d + retail_d +
    garage_d + storage_d + factory_d + other_d, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-1069.09  -238.48    91.06   239.74   801.84

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.784e+02  1.014e+01  57.045  < 2e-16 ***
aadt        -1.774e-03  1.037e-04 -17.107  < 2e-16 ***
crash        2.499e+00  1.167e-01  21.417  < 2e-16 ***
res_d        5.692e+01  9.682e+00   5.879 4.42e-09 ***
office_d    -2.180e+01  8.709e+00  -2.503 0.012353 *
retail_d     1.048e+02  7.907e+01   1.325 0.185214
garage_d    -1.098e+02  1.358e+02  -0.808 0.418891
storage_d    1.164e+02  1.580e+02   0.736 0.461567
factory_d    5.661e+02  1.500e+02   3.774 0.000163 ***
other_d      3.703e+01  4.698e+01   0.788 0.430626
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 317.7 on 4449 degrees of freedom
Multiple R-squared:  0.1729,    Adjusted R-squared:  0.1713
F-statistic: 103.4 on 9 and 4449 DF,  p-value: < 2.2e-16
```

Part VI. Dashboard

- To help visualize NPMRDS, a web-based dashboard was created by using R shiny, leaflet, echarts4r, html and css which displays both county & segment level congestion using gauge, histogram, line graph, heatmap, calendar and spatial map.

- As required by federal government, Congestion Management Process requires MPO to make data-driven decisions on combating congestions. This dashboard can be used as a data tool as well as reporting, which helps NYMTC commit the federal requirement.

Access Link:

https://yazhang.shinyapps.io/dash/

Thank you!

yaxinzhang022@hotmail.com

Question?