Transportation Data Visualization

CEE 412/ CET 522

Transportation Data Management and Visualization

WINTER 2020

Announcement

Project 2 (Final Project) will release today.

A portion of today's lecture content comes from CSE 512 (https://courses.cs.washington.edu/courses/cse512/19sp/). Thanks to Prof. Jeffrey Heer, who shares the lecture slides of CSE 512 Data Visualization.

Outline

Value of Data Visualization

Data Visualization Techniques

Visualization Tools

Transportation Data Visualization

Value of Visualization

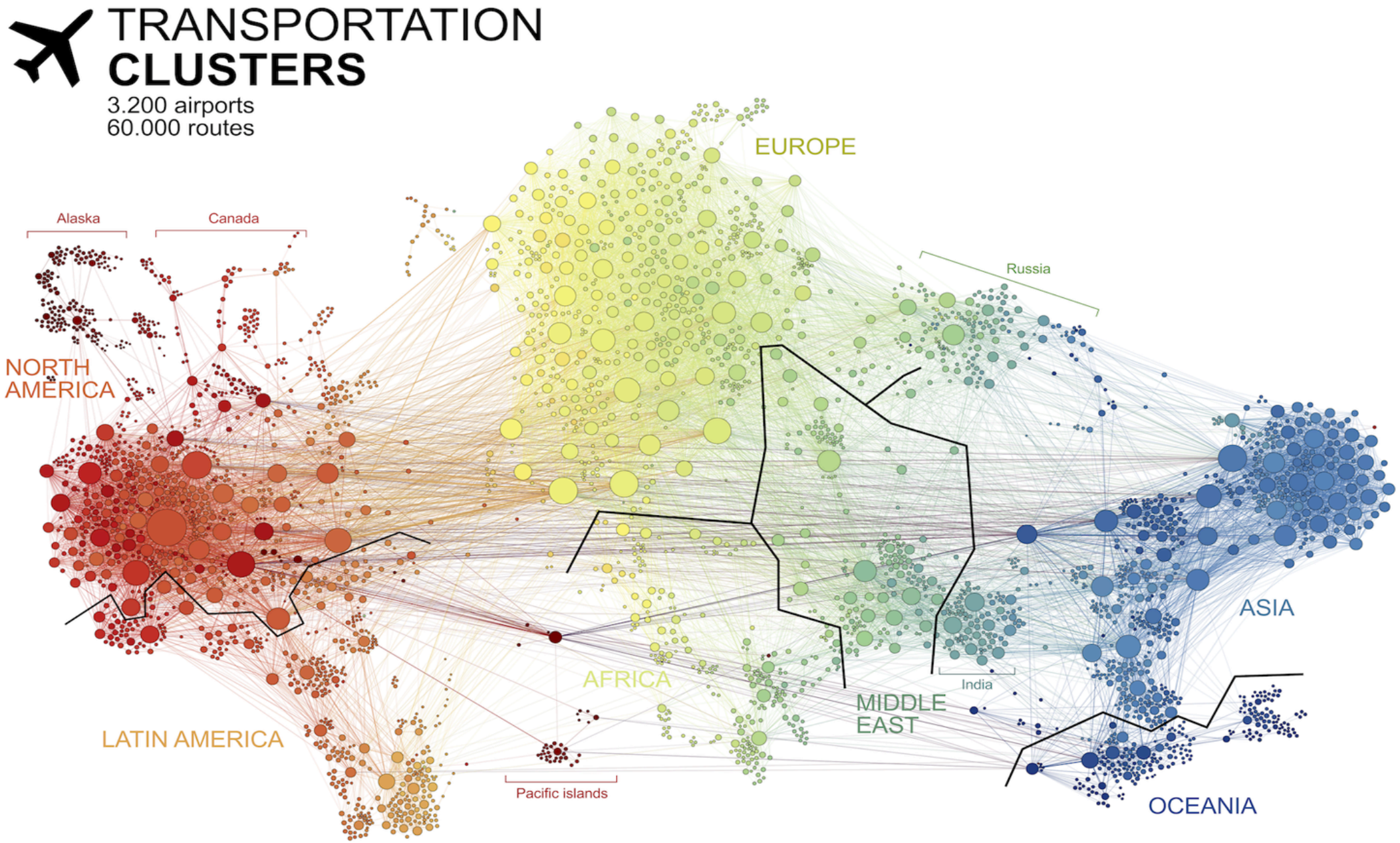
https://3blaws.s3.amazonaws.com/images/SmartTrans_for_3BL.jpg

What is Visualization?

"Transformation of the symbolic into the geometric" [McCormick et al. 1987]

"... finding the artificial memory that best supports our natural means of perception." [Bertin 1967]

"The use of computer-generated, interactive, visual representations of data to amplify cognition." [Card, Mackinlay, & Shneiderman 1999]



https://www.oreilly.com/library/view/graph-algorithms/9781492047674/assets/gral_0103.png

Why Create Visualization?

Answer questions (or discover them)

Make decisions

See data in context

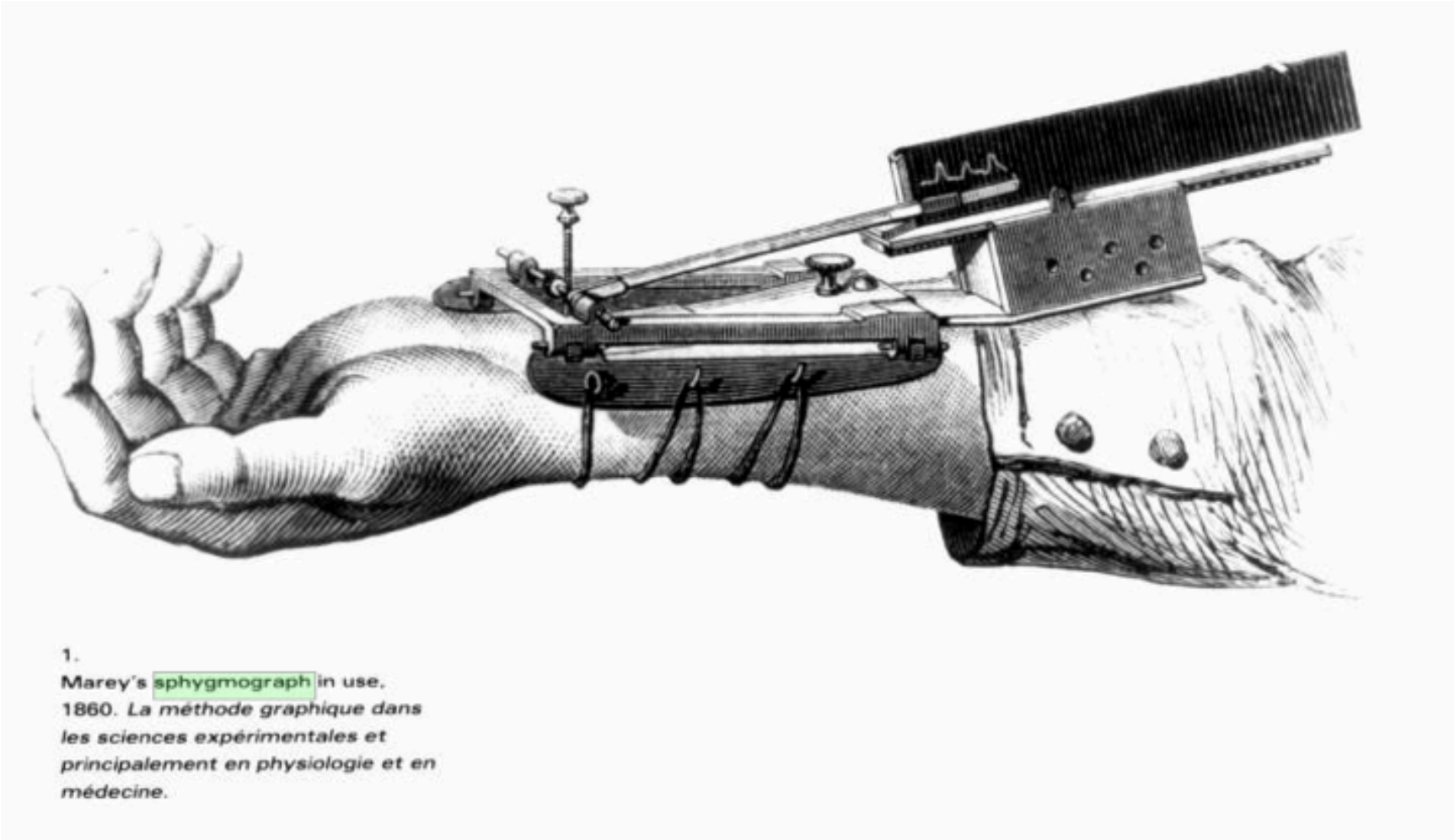
Support graphical calculation

Find patterns

Present argument or tell a story

Inspire

Record Information

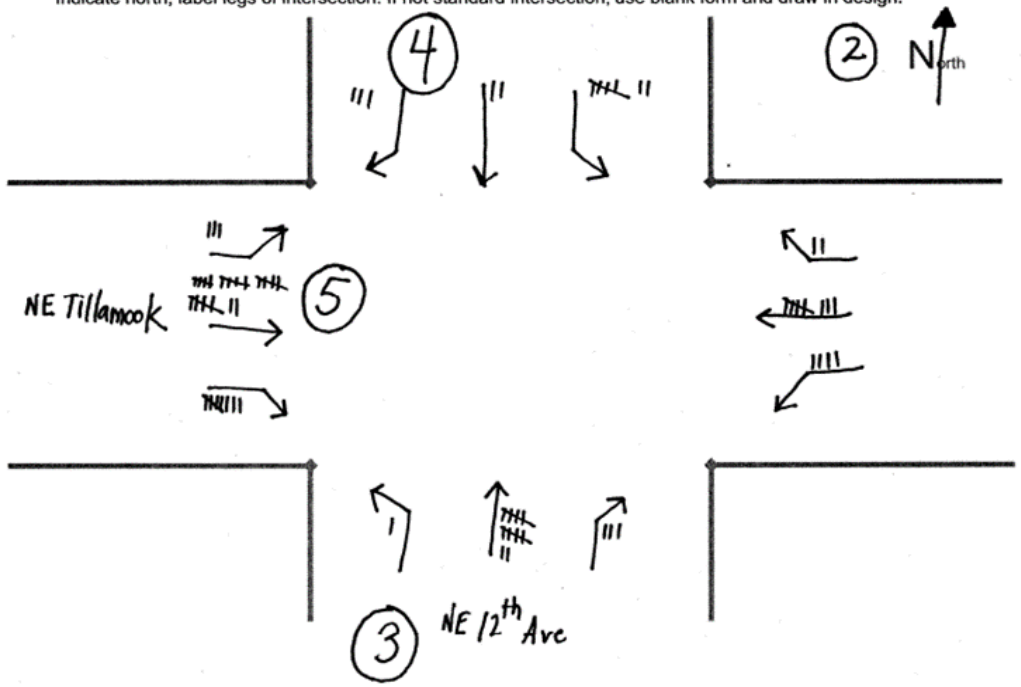


1.
Marey's sphygmograph in use, 1860. *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine.*

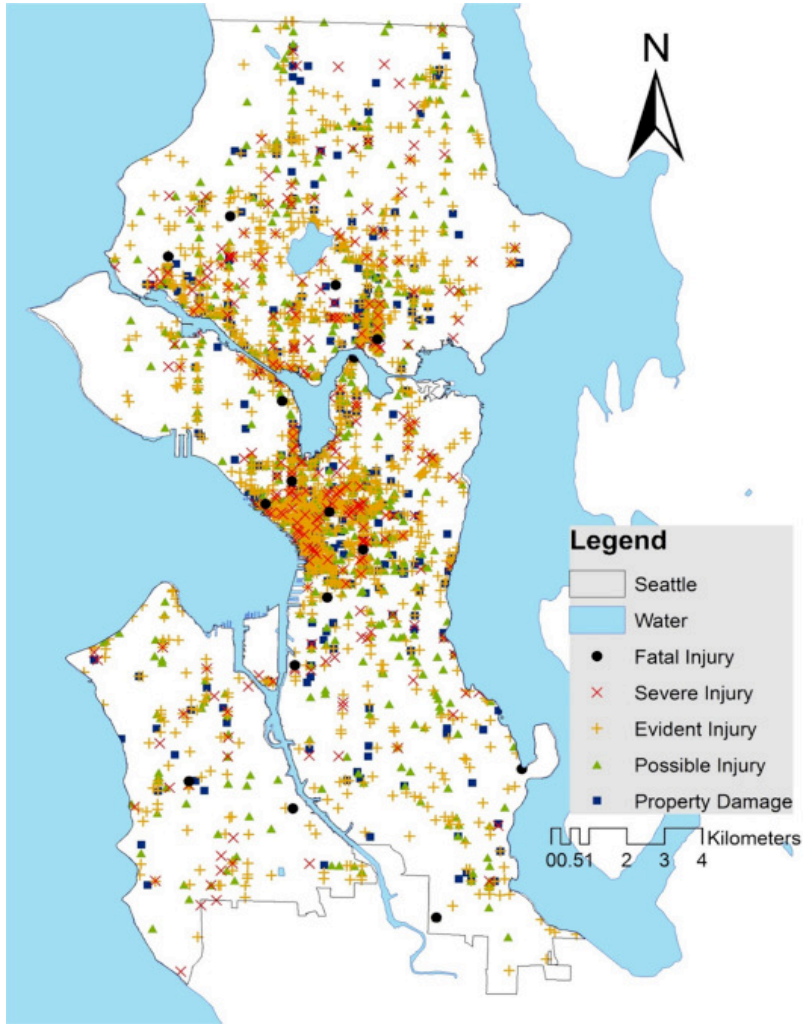
Record Information

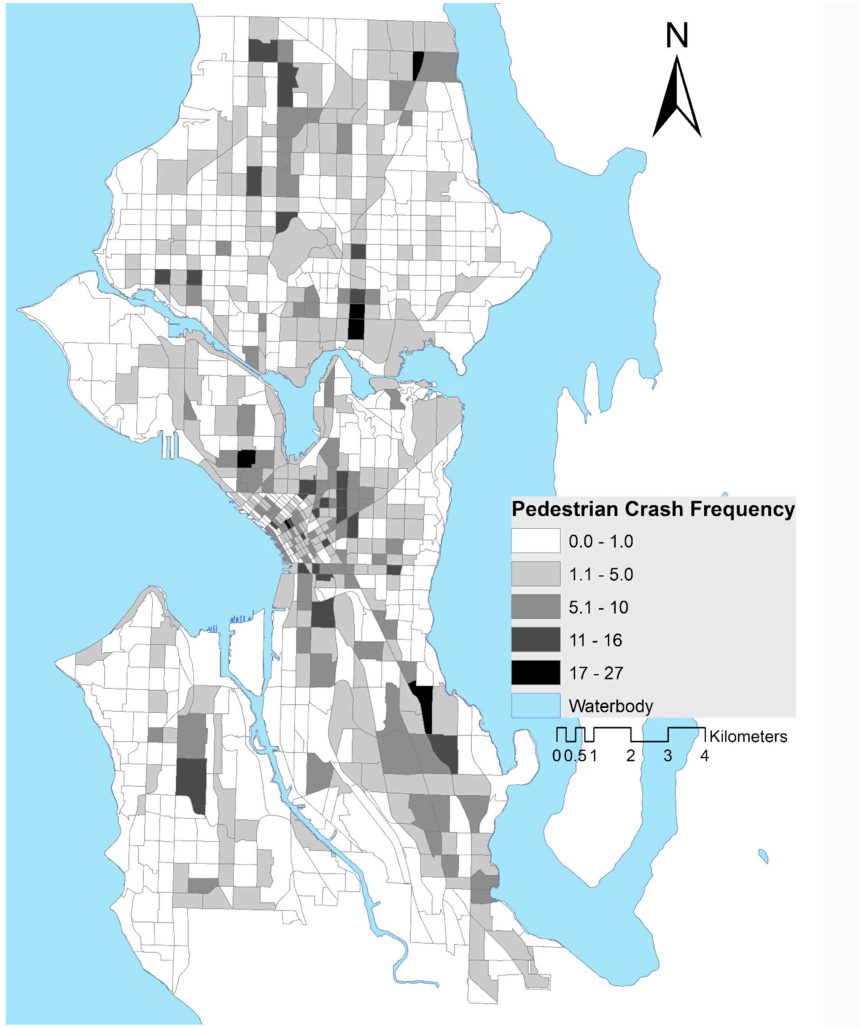
	Male	Female	Total
with helmet	5		59
without helmet			16
totals →	6 45	30	75 7

Indicate movement thru intersection by using arrows to show direction of travel (left turns, through, etc.).
Indicate north, label legs of intersection. If not standard intersection, use blank form and draw in design.



Support Reasoning





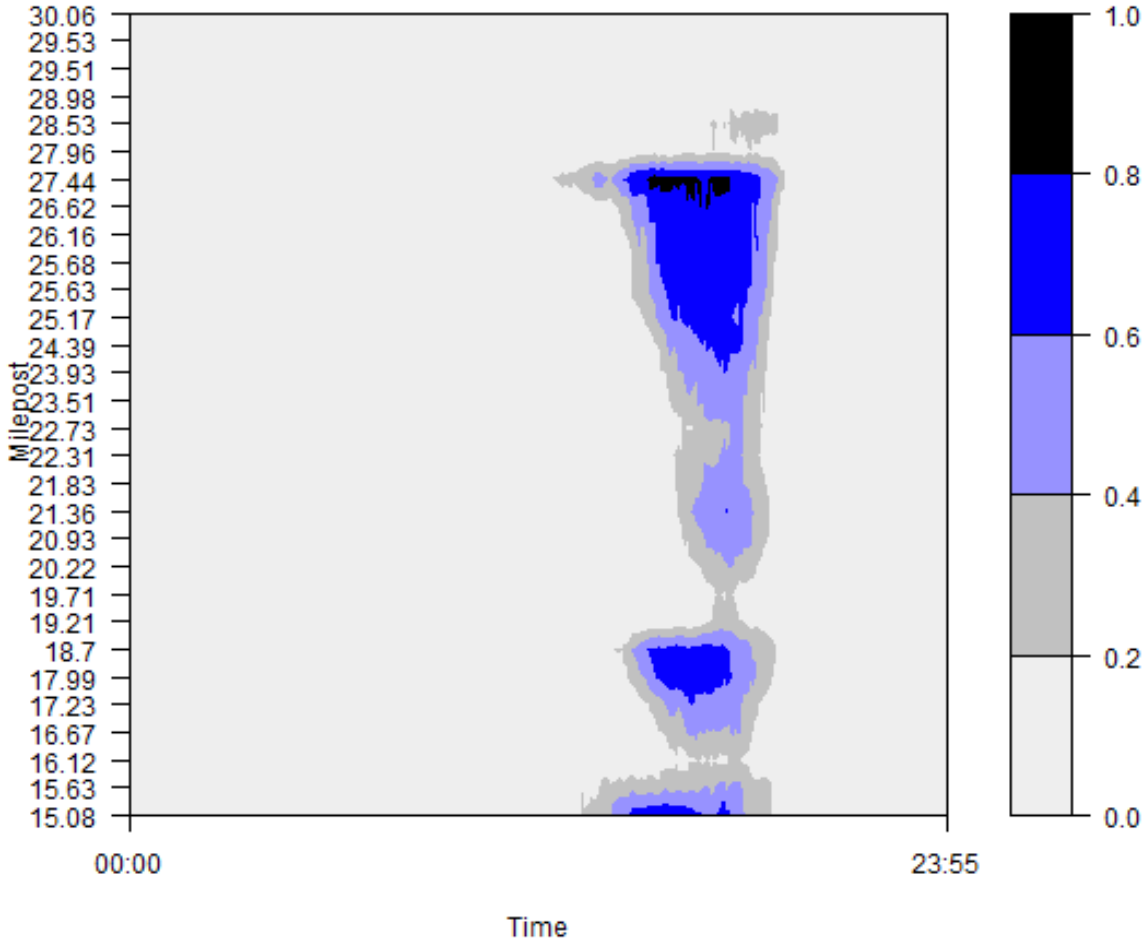
https://ars.els-cdn.com/content/image/1-s2.0-S2214140516301864-gr2.jpg

https://ars.els-cdn.com/content/image/1-s2.0-S0001457515301184-gr1.jpg

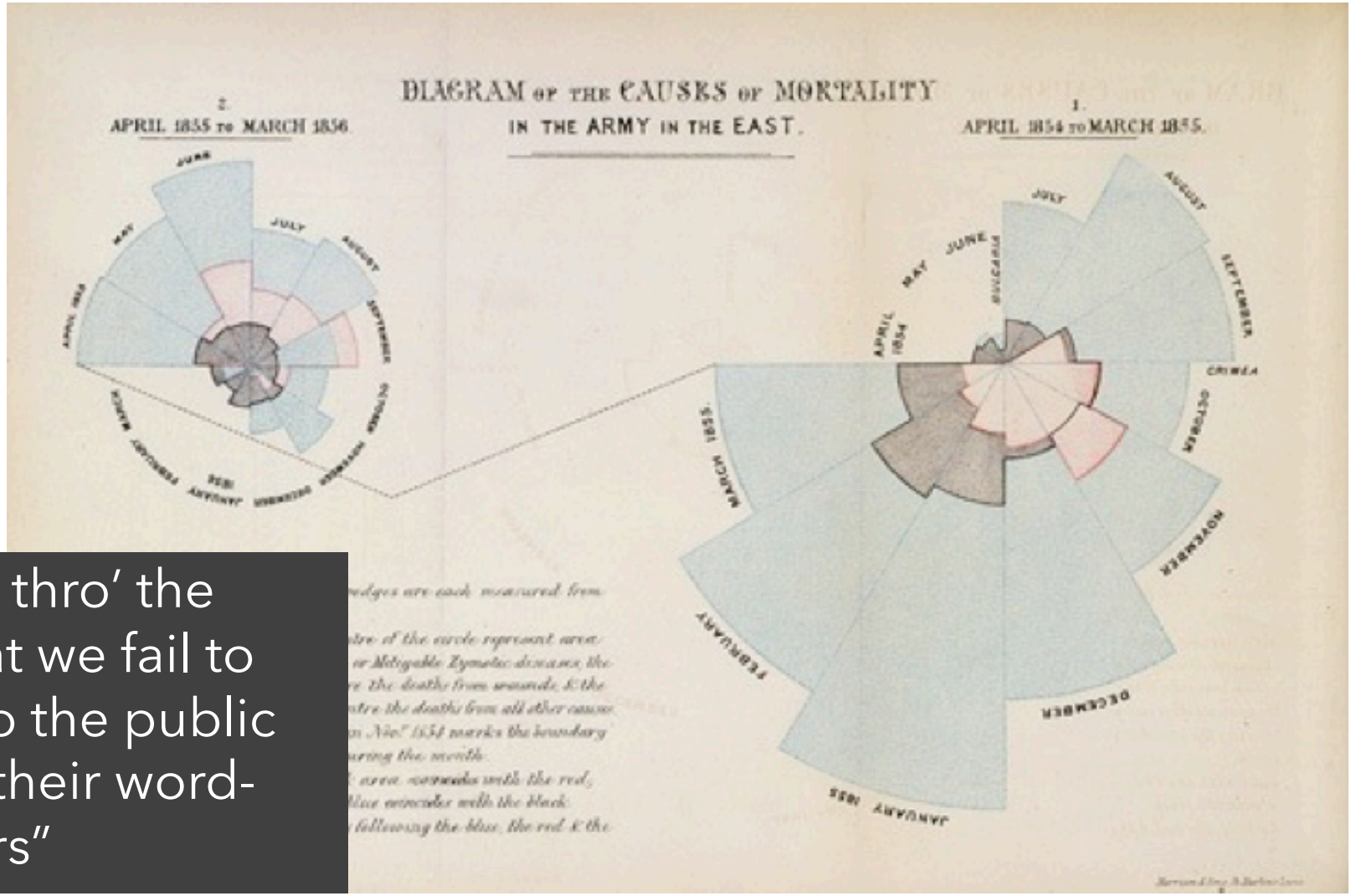
Support Reasoning

Congestion Plot

- Route 405



Convey Information to Others

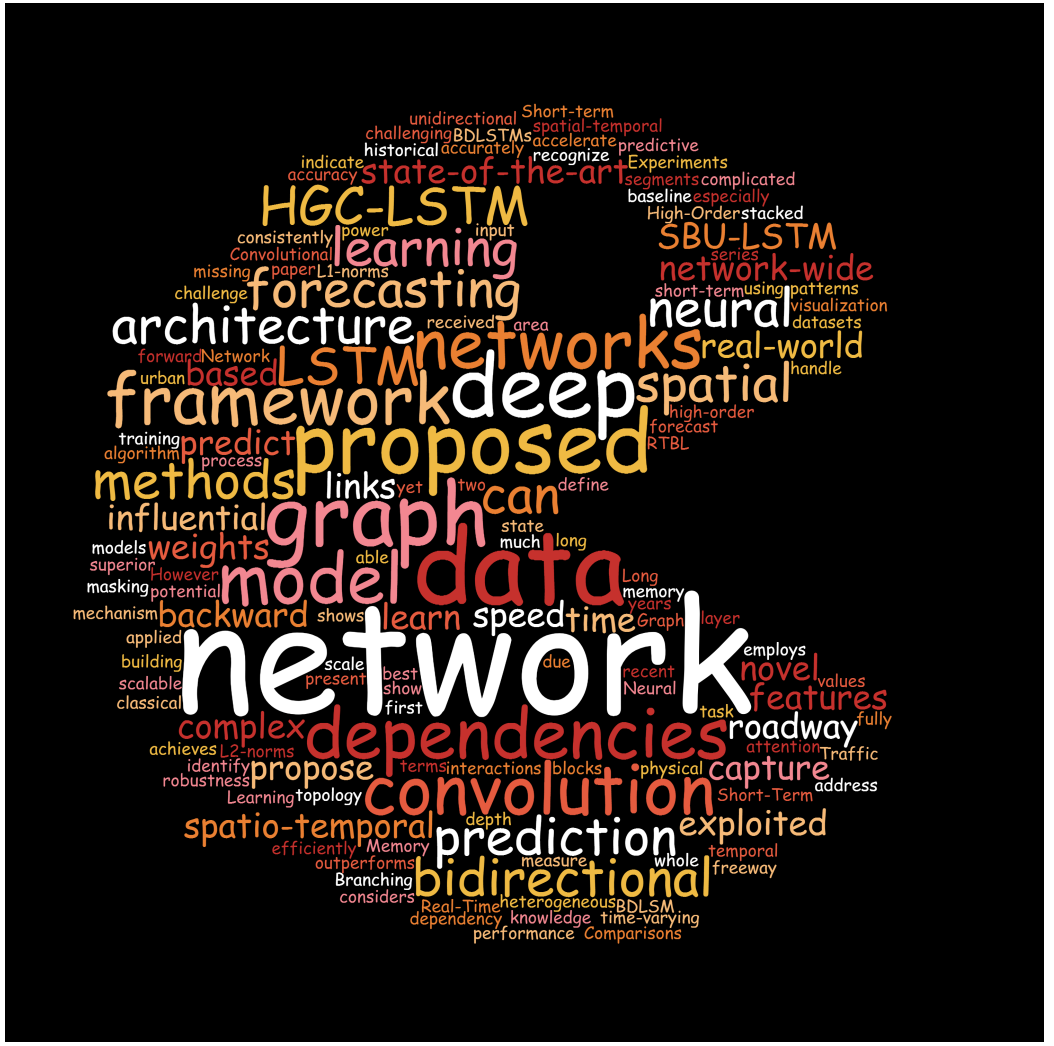


"to affect thro' the Eyes what we fail to convey to the public through their word-proof ears"

Convey Information to Others

Word Cloud

- Summarize a text



The value of Visualization

Record information

- Blueprints, photographs, seismographs, …

Analyze data to support reasoning

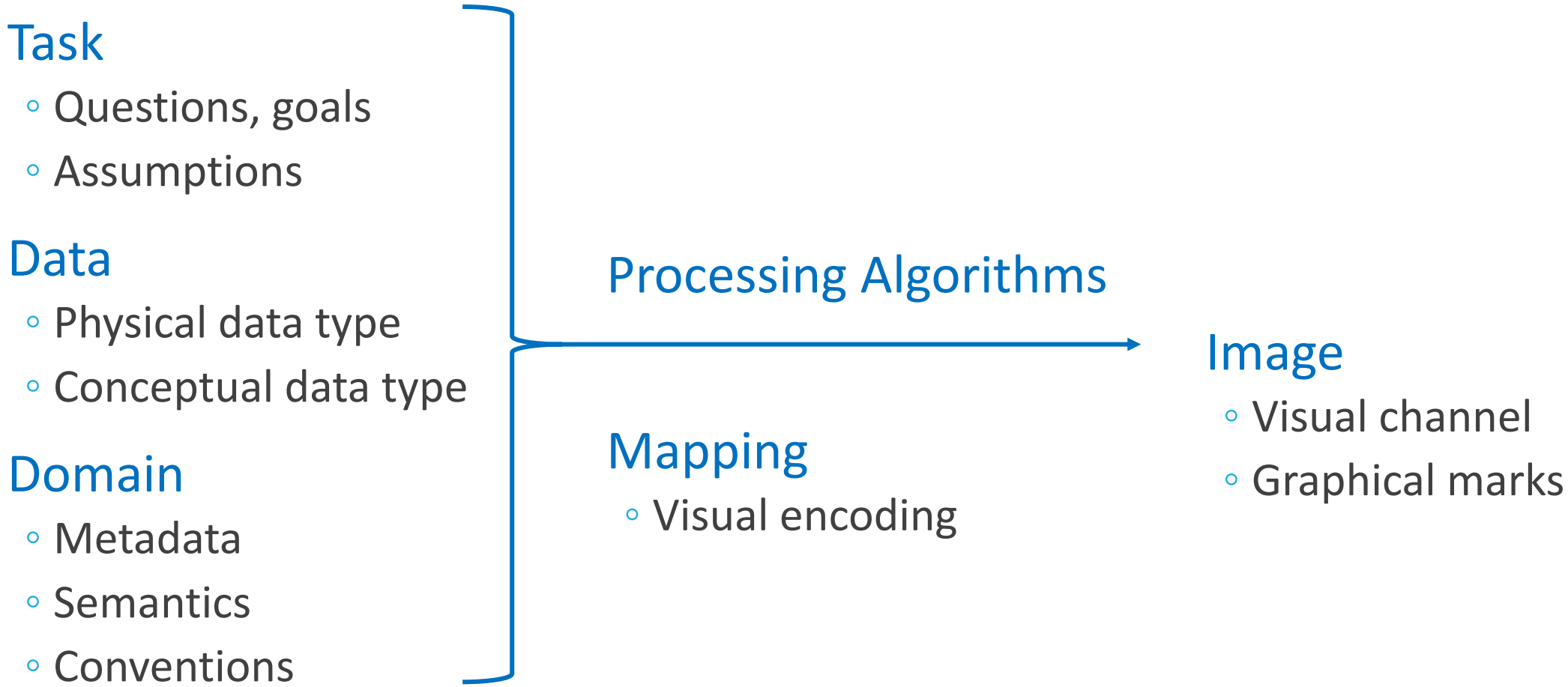
- Develop and assess hypotheses
- Find patterns / Discover errors in data
- Expand memory

Communicate information to others

- Share and persuade
- Collaborate and revise

Data Visualization Techniques

The Big Picture



Data Types for Visualization

1D (sets and sequences)

Temporal

2D (maps)

3D (shapes)

nD (relational)

Trees (hierarchies)

Networks (graphs)

Spatiotemporal (trajectories)

Are there others?

Types of Visualization

Scatter Plot

Line Chart

Pie Chart

Histogram

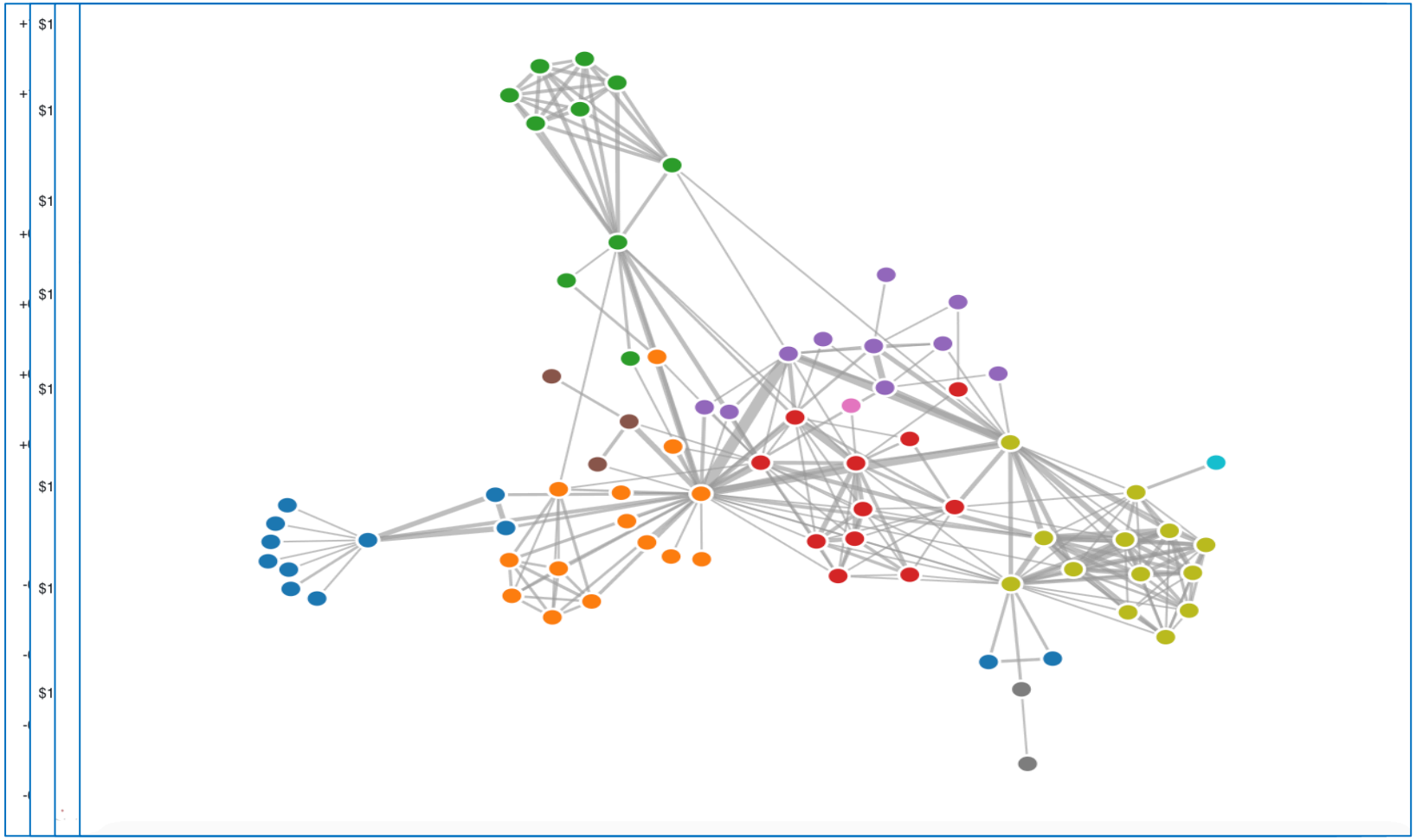
Heatmap

Tree

Map

Graph

What else ?



Visual Encoding Design

A Design Space of Visual Encodings

- Use **expressive** and **effective** encodings
- Avoid **over-encoding**
- **Reduce** the problem space
- Use **space** and **small multiples** intelligently
- Use **interaction** to generate *relevant* views

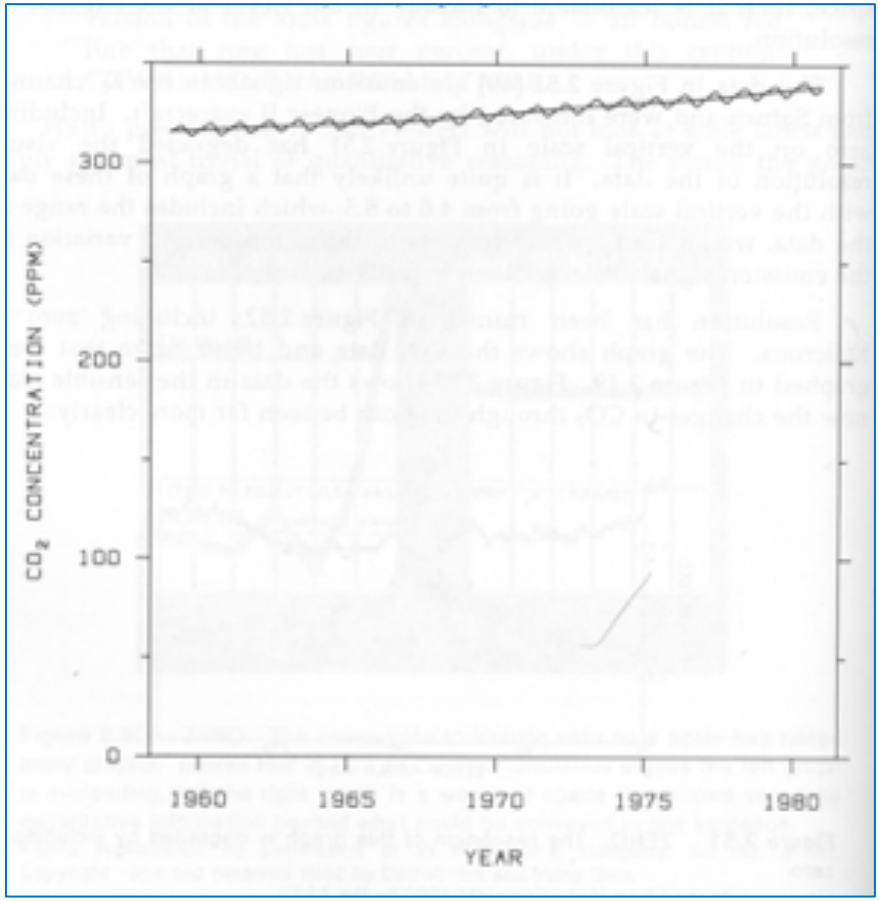
Rarely does a single visualization answer all questions. Instead, the ability to generate appropriate visualizations quickly is critical!

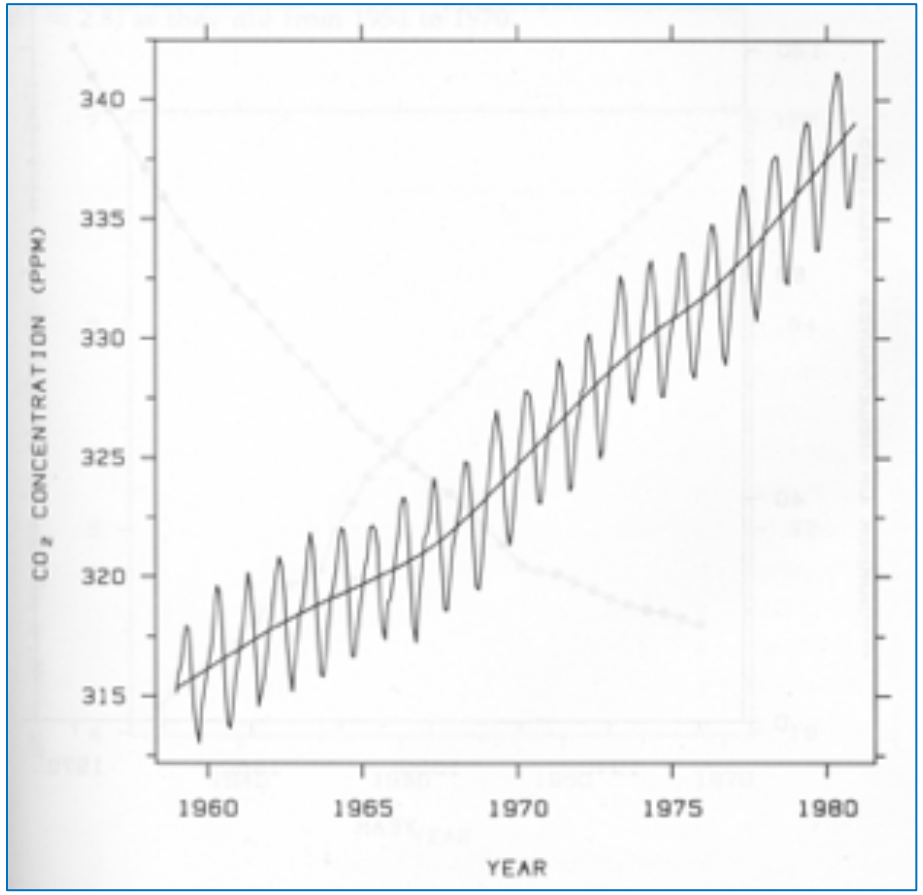
Encoding Design Effectiveness

Several aspects of effective design that should be noted:

- Scales & Axes
- Aspect Ratio
- Multidimensional Data
- Parallel Coordinates
- Dimensionality Reduction

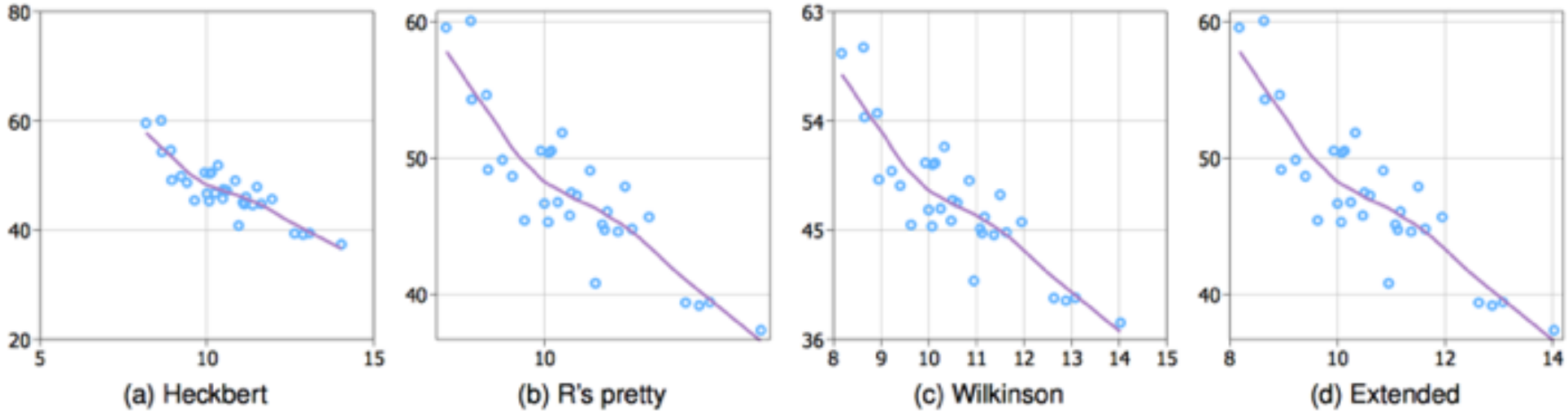
Encoding Effectiveness - Scales & Axes





Yearly CO_2 concentrations [Cleveland 85]

Encoding Effectiveness - Scales & Axes



Simplicity - numbers are multiples of 10, 5, 2

Coverage - ticks near the ends of the data

Density - not too many, nor too few

Legibility - whitespace, horizontal text, size

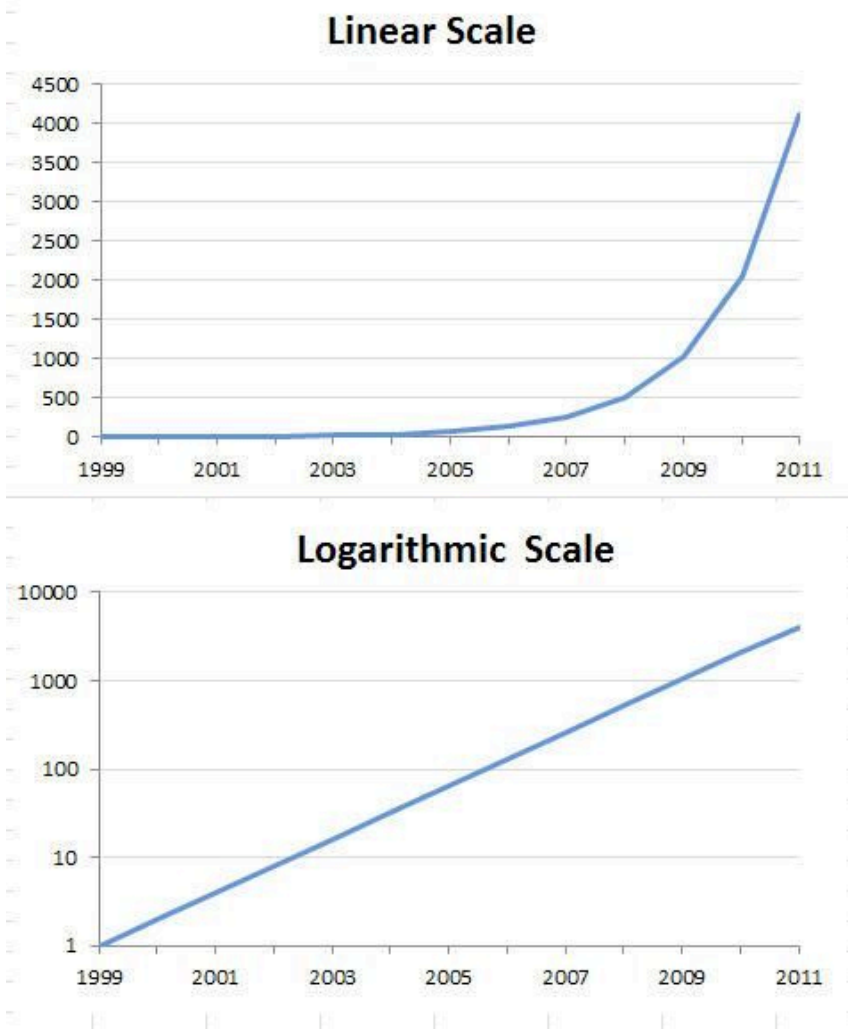
Encoding Effectiveness - Scales & Axes

Linear Scale vs. Log Scale

- **Address data skew** (e.g., long tails, outliers)
- Enables comparison within and across multiple orders of magnitude.
- **Focus on multiplicative factors** (not additive)
- Recall that the logarithm transforms **×** to **+**!
- Percentage change, not linear difference.

Constraint: **positive, non-zero values**

Constraint: **audience familiarity?**

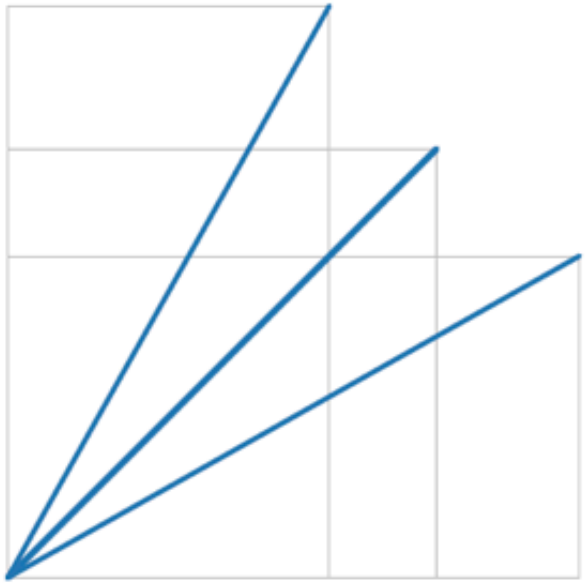


https://thumbor.forbes.com/thumbor/474x586/https://blogs-images.forbes.com/naomirobbins/files/2012/01/linear_log.jpg?width=960

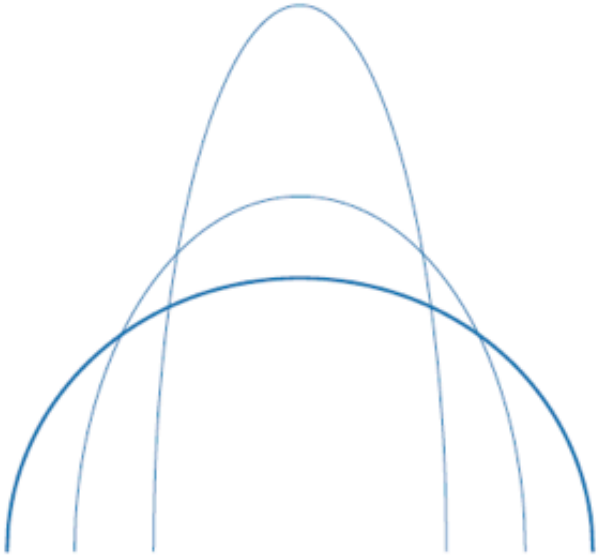
Encoding Effectiveness - Aspect Ratio

Banking to 45°

To facilitate perception of trends, maximize the discriminability of line segment orientations



Straight line -> 45°

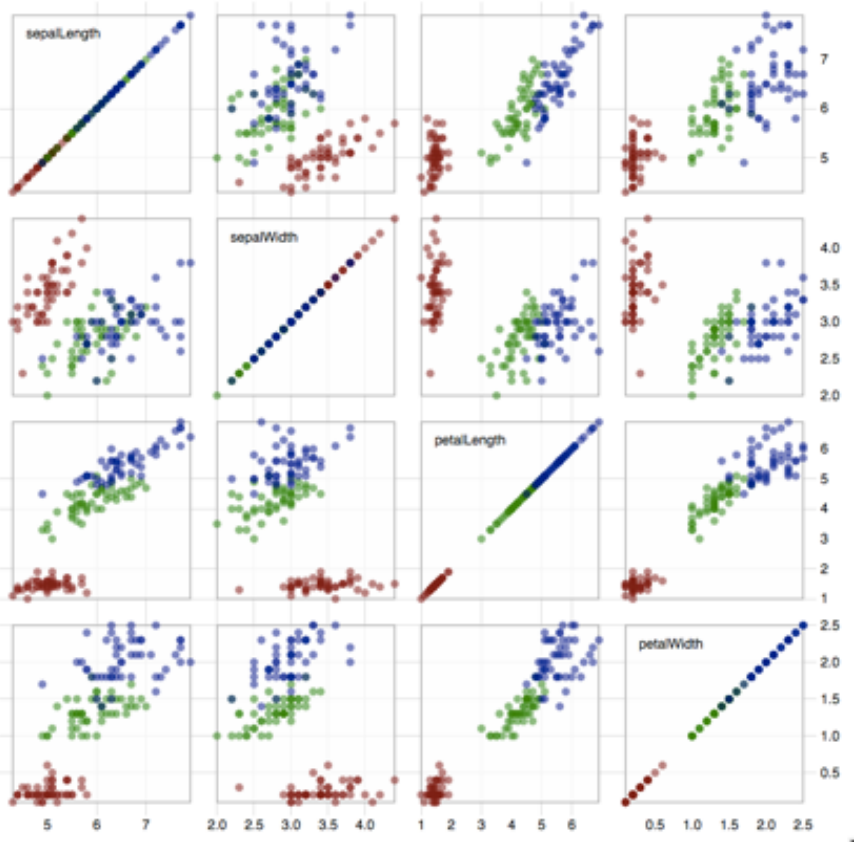


Ellipse -> Circle

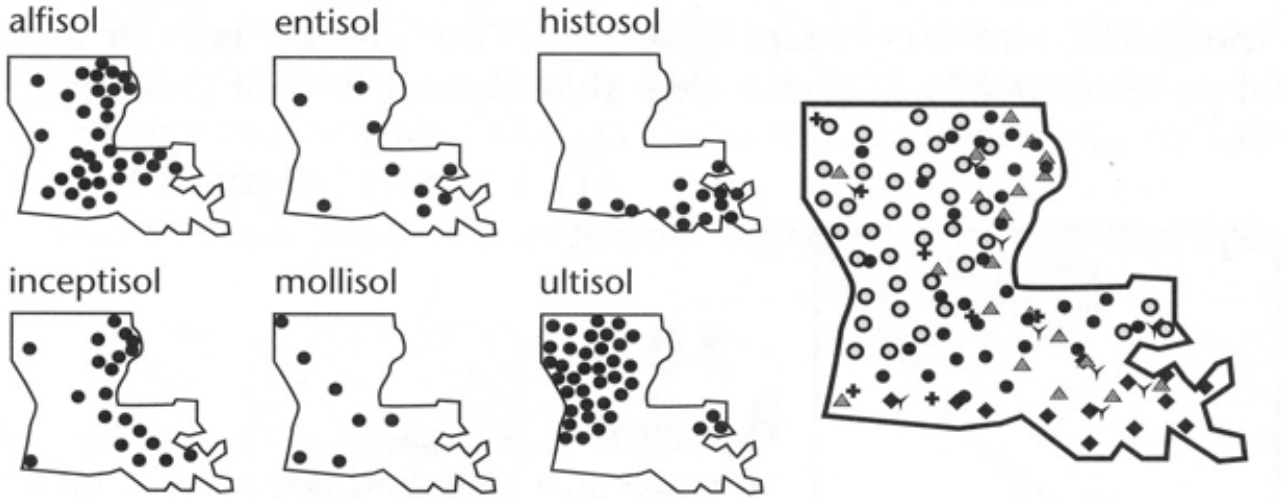
[Talbot et al. 2011]

Encoding Effectiveness - Multidimensional Data

Scatterplot Matrix



Small Multiples



[MacEachren '95, Figure 2.11, p. 38]

Encoding Effectiveness - Parallel Coordinates

1. Draw parallel axes for each variable

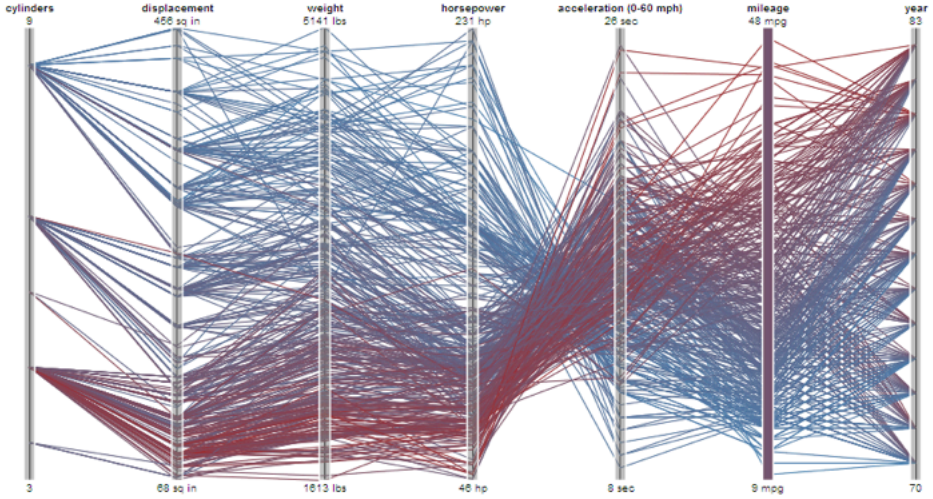
2. For each tuple, connect points on each axis

Between adjacent axes: line crossings imply neg. correlation, shared slopes imply pos. correlation.

Full plot can be cluttered. **Interactive selection** can be used to assess multivariate relationships.

Highly sensitive to axis **scale** and **ordering**.

Expertise required to use effectively!



Encoding Effectiveness - Dimensionality Reduction

Many Reduction Techniques!

- Principal Components Analysis (PCA)
- t-Dist. Stochastic Neighbor Embedding (t-SNE)
- Uniform Manifold Approx. & Projection (UMAP)
- Auto-Encoder Neural Networks
- Multi-dimensional Scaling (MDS)
- …

Interaction

Interaction between people and machines requires *mutual intelligibility* or *shared understanding*.

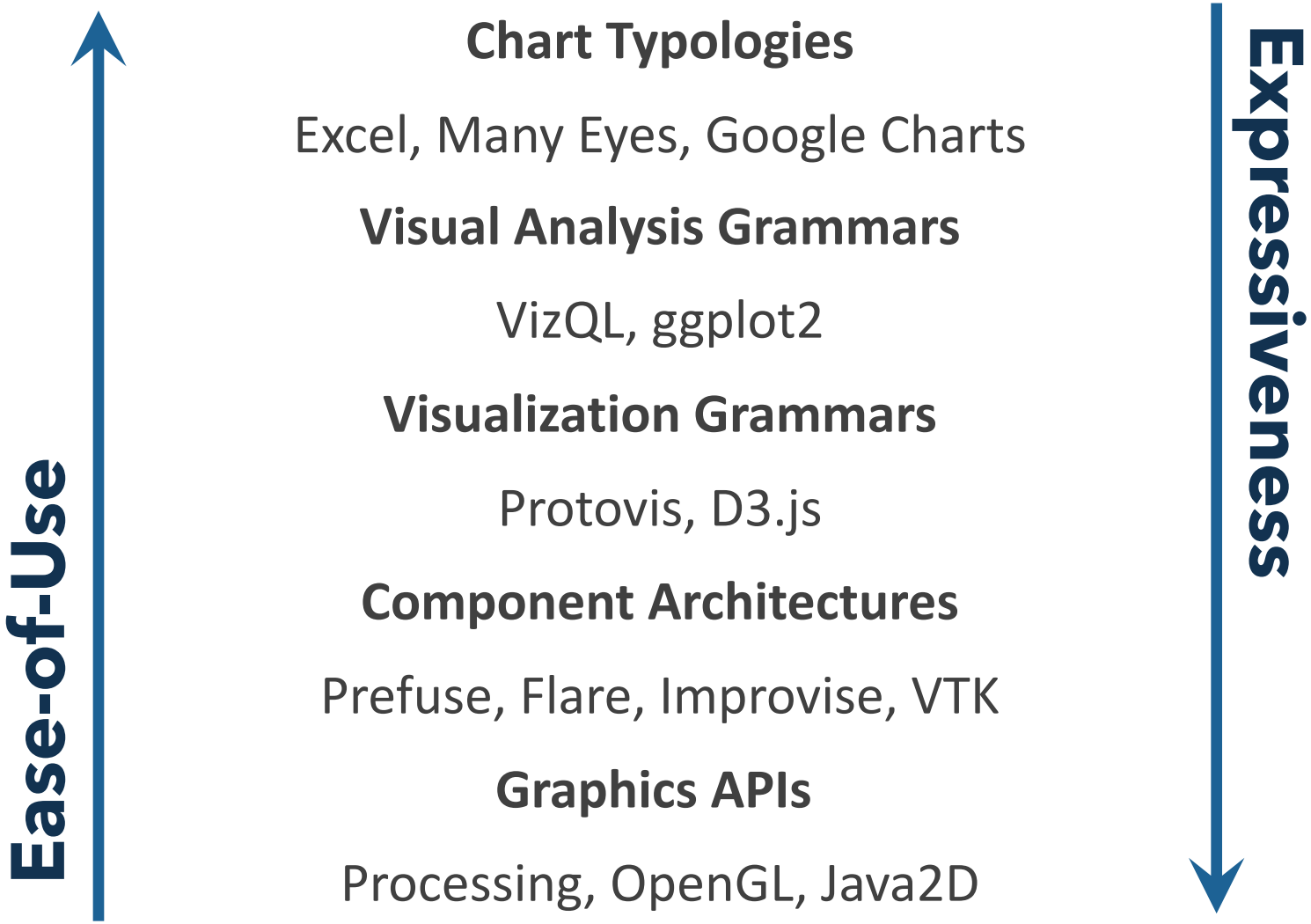
Taxonomy/Classification of Interactions

- **Data and View Specification**
  - Visualize, Filter, Sort, Derive
- **View Manipulation**
  - Select, Navigate, Coordinate, Organize
- **Process and Provenance**
  - Record, Annotate, Share, Guide

Interactive Visualization, or even animation, is useful for data visualization tools or data analysis platforms

Visualization Tools

Visualization Tools



Visualization Tools

ggplot2 – R (https://ggplot2.tidyverse.org/)

- ggplot2 is a data visualization package for the statistical programming language R. Created by Hadley Wickham in 2005
- ggplot2 is an implementation of Leland Wilkinson's Grammar of Graphics—a general scheme for data visualization which breaks up graphs into semantic components such as scales and layers.

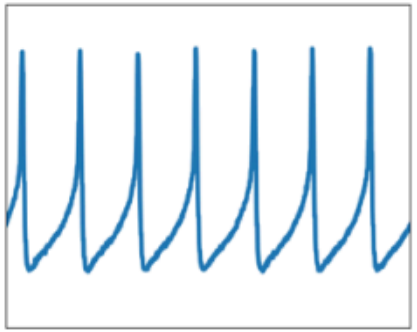
Cheatsheets (https://github.com/rstudio/cheatsheets/blob/master/data-visualization-2.1.pdf)

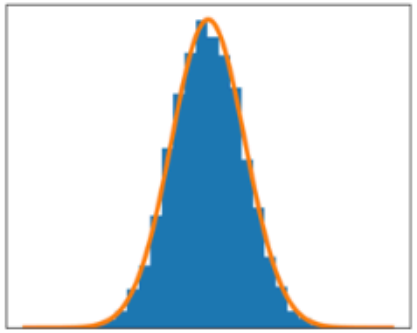
```
install.packages("ggplot2")
library(ggplot2)
```

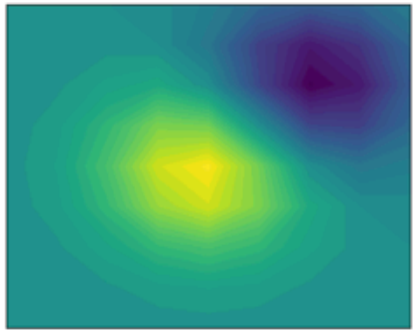
Visualization Tools

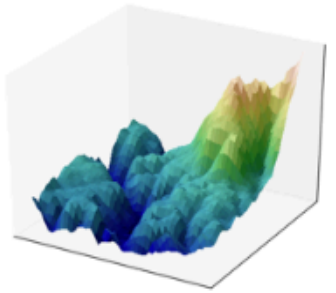
Matplotlib – Python (https://matplotlib.org/)

- Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.
- Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.





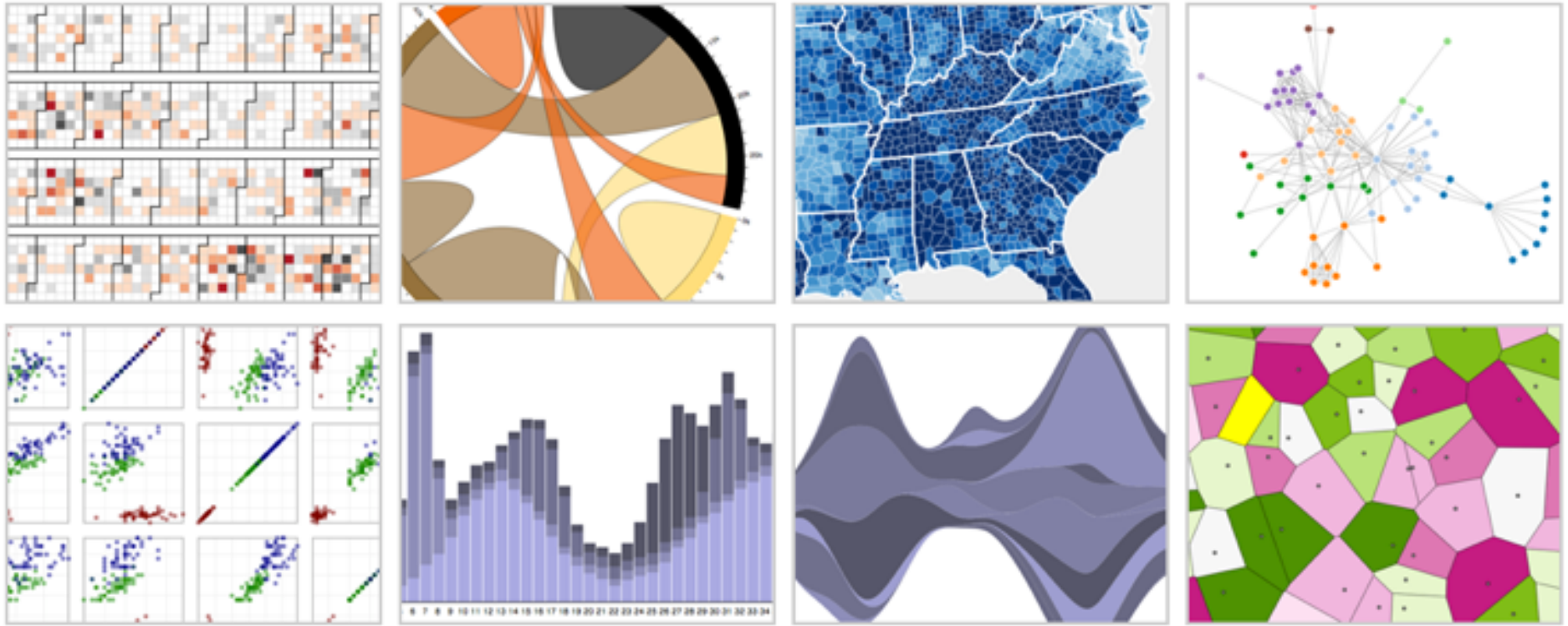




matplotlib

Visualization Tools

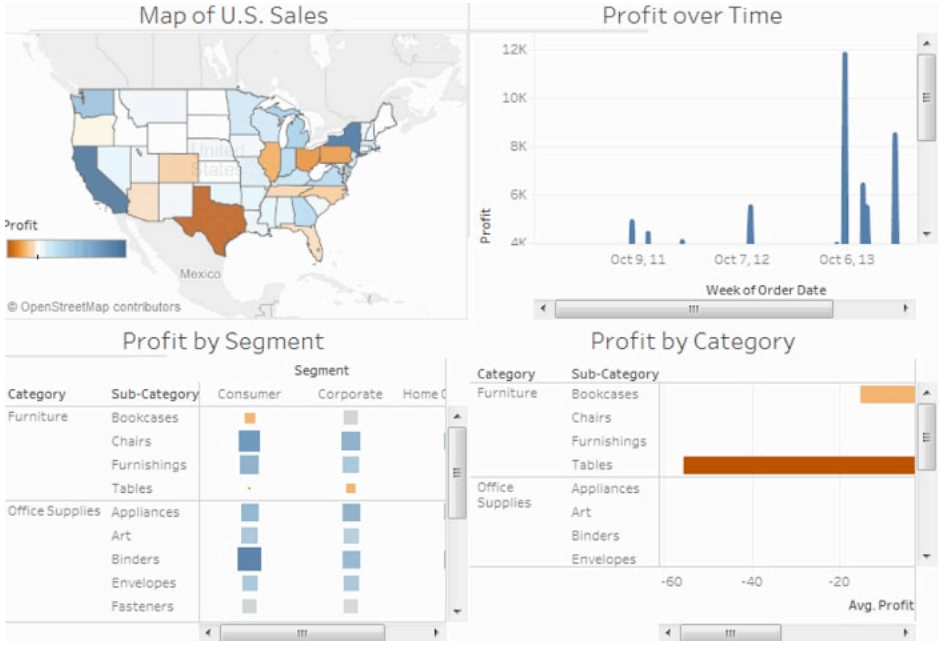
d3.js – JavaScript (https://d3js.org/)



Tableau

Tableau (https://www.tableau.com/)

- A powerful and fastest growing data visualization **tool** used in the Business Intelligence Industry.
- It helps in simplifying raw data into the very easily understandable format.
- Data analysis is very fast with **Tableau** and the visualizations created are in the form of dashboards and worksheets.



- **Tableau** Public is **free**
- One of the most most popular
- **Not** difficult to learn
- Not covered in this class

https://www.kdnuggets.com/wp-content/uploads/tableau-crowded-dashboard.jpg

Visualization Tools

Useful Resources in CSE 512 Data Visualization

- Tools
  - Visualization Toolkits: D3, Vega, Leaflet …
  - Visualization Tools: Tableau, ggplot2 …
  - Network Analysis Tools: Gephi, NetworkX …
  - Color tools

- Tutorials & Tips
  - HTML, CSS
  - JavaScript
  - d3.js
  - Git & GitHub

Transportation Data Visualization

Visualization for Transportation Data

Depends on Requirements of Tasks

- Safety analysis
- Traffic signal control
- Traffic assignment

Depends on Data Types

- Tabular data: relational database, CSV…
- Geospatial data: coordinates, geometric data
- Trajectory Data

Depends on Targets of Tasks

- Visualize statistical models: various charts
- Visualize geospatial info
- …

Visualization for Transportation Data

Data Viewing

Data manipulation/processing

Data modeling/analysis

Analysis results interpretation

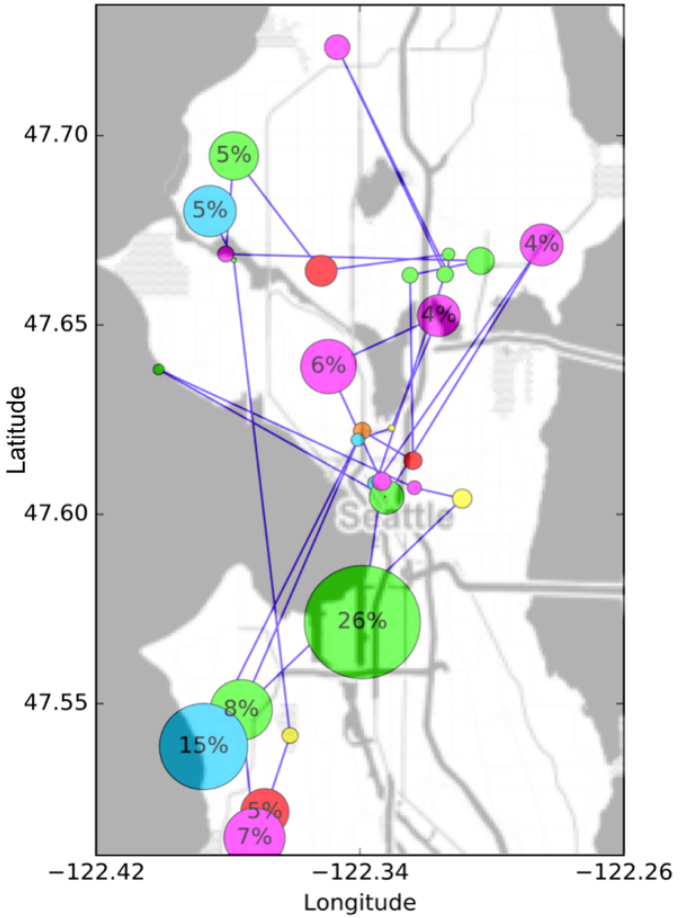
Taking a research paper/report for an example, it contains several sections, including introduction, literature review, methodology, experiments, conclusion. Normally, data visualization can (need to) be presented in all sections except for conclusion.

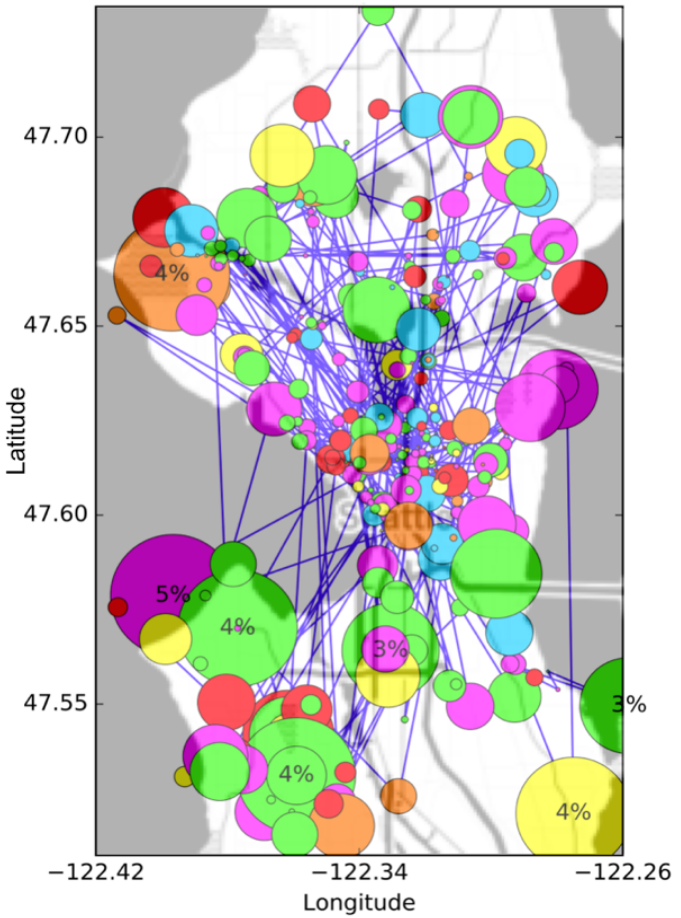
Visualization in Data Viewing

Car2Go Origin-Destination (OD) Data

Two Vehicles' OD over time

- Line: from O to D
- Circle size: unoccupied time
- If we have the fuel status of the vehicle, what kind of component can help us to visualize that info?



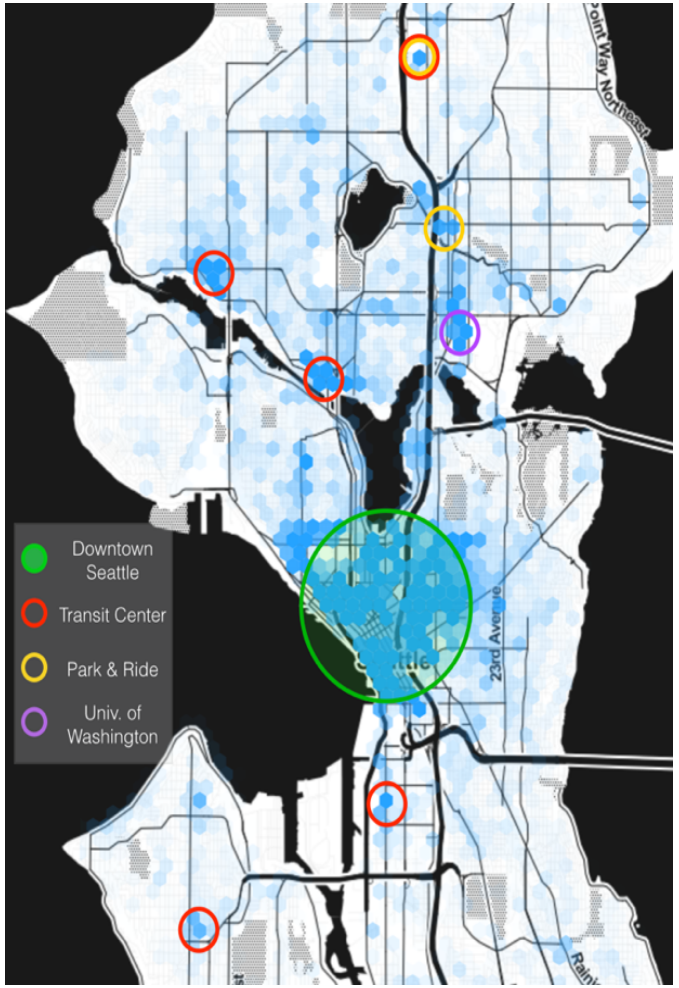


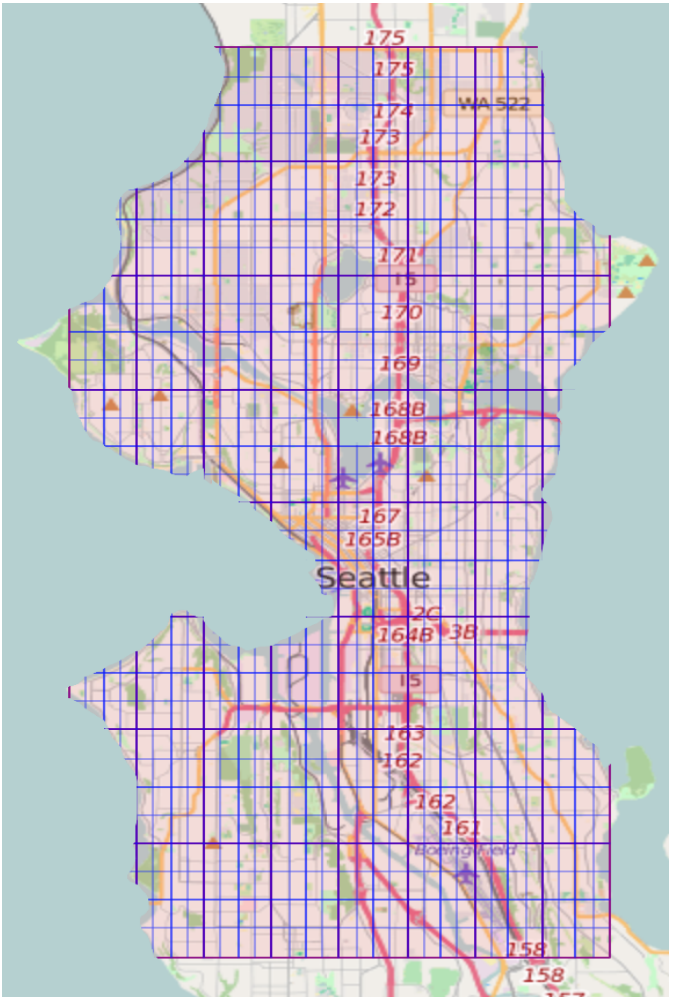
Visualization in Data Processing

Car2Go OD distribution

- Split into grids
- Split into hexagonal tiling

Facilitate to figure out the hotspots

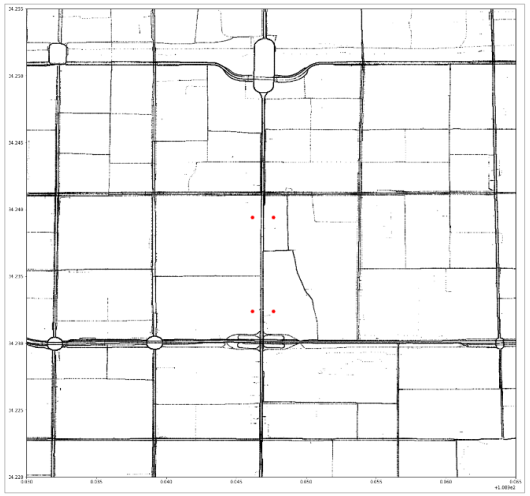


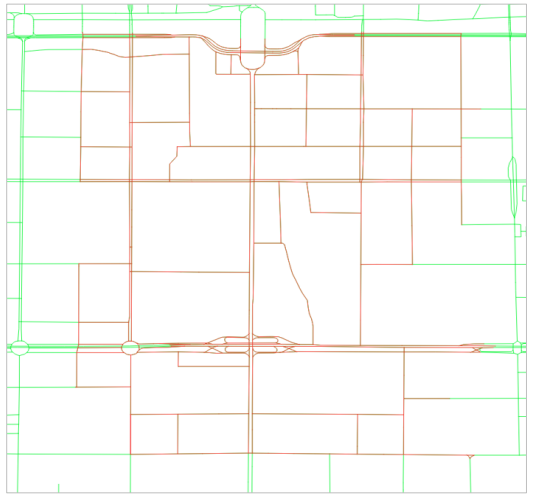


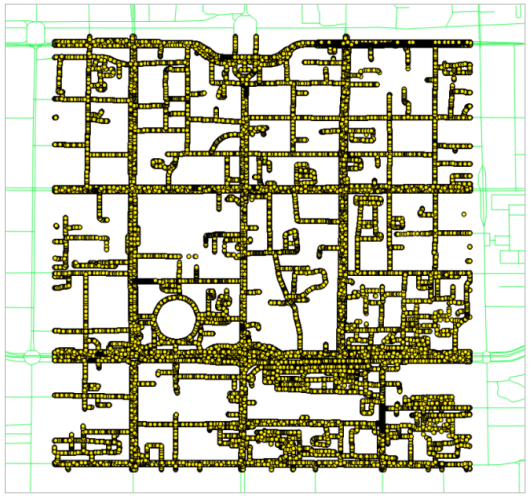
Visualization in Data Processing

Didi trajectory data

- Target: traffic forecasting
- Procedures:
  - Trajectory data → mapping GPS points to road segment → road segment based speed/travel time data
- Tools: QGIS, OpenStreetMap, Python

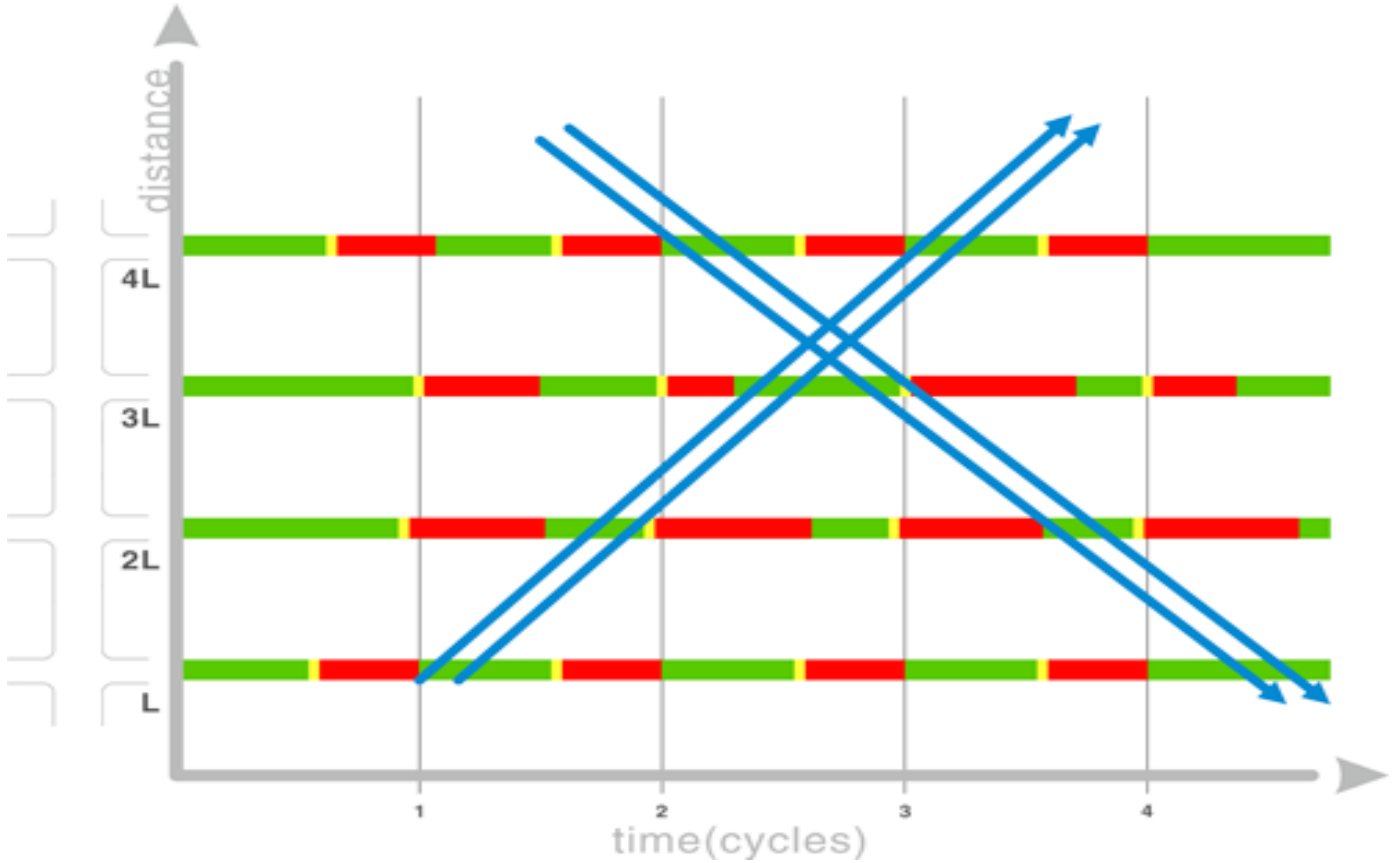






Visualization in Data Modeling

Coordinated Traffic Light Control



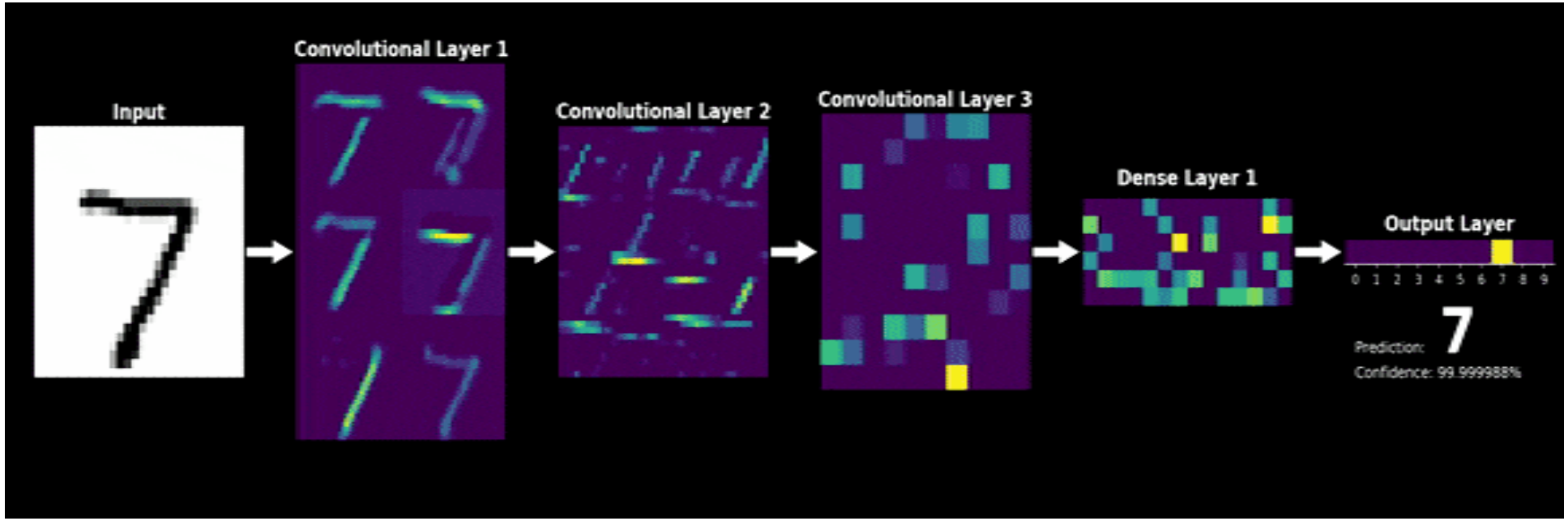
https://ops.fhwa.dot.gov/publications/fhwahop08024/images/6_1.png

Visualization in Data Modeling

Visualization during the modeling

- Parameters of a model
- Errors of fitted model

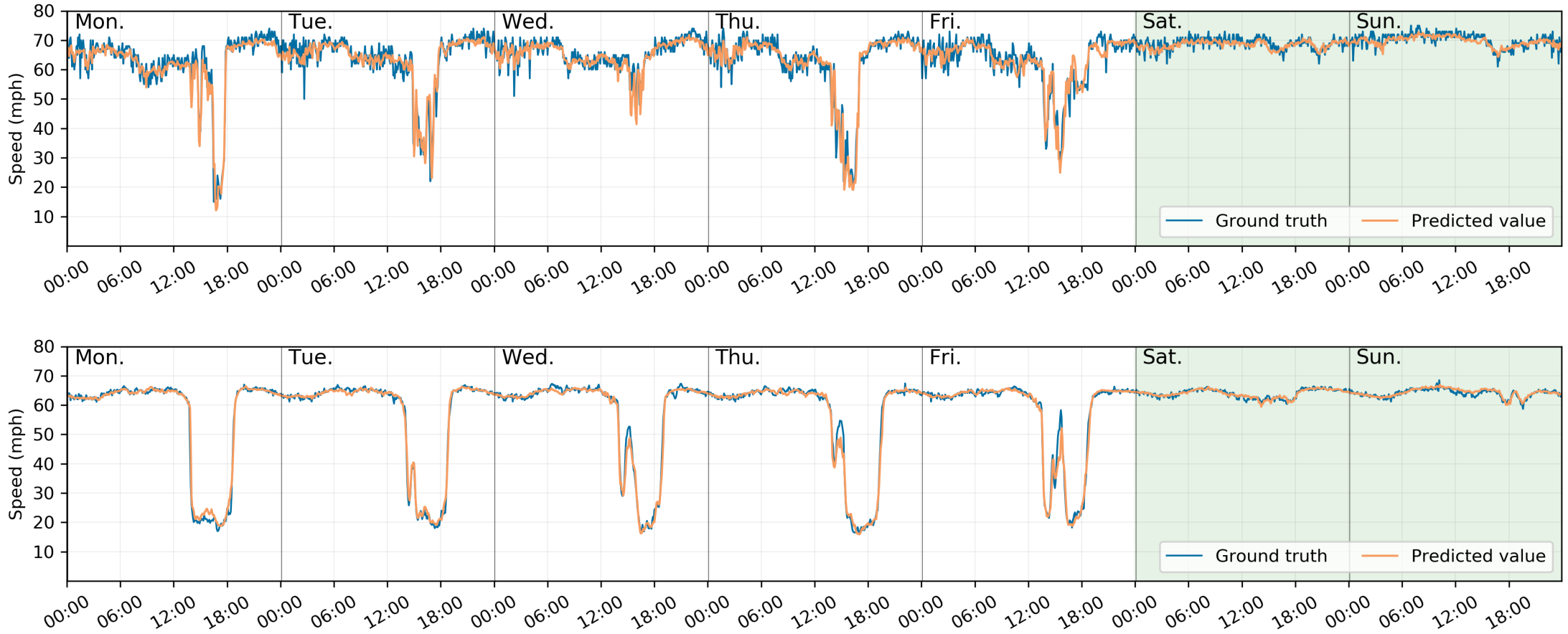
Assist to adjust model parameters or refine model design



https://external-preview.redd.it/Dhrpp8M4X9BpyFOKGpD6uxl2aFRC3fBS-akgcZ2cxYw.gif?format=png8&s=8a9143099e235e11018e7adcadcb7b7973f5e4c1

Visualization in Results Interpretation

Traffic Speed Prediction Results



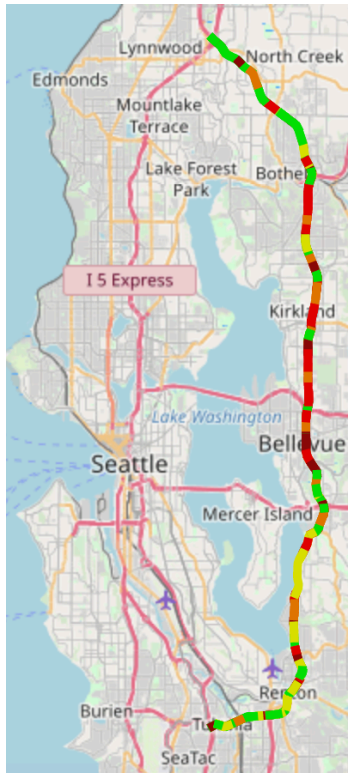
Visualization in Results Interpretation

Traffic accident analysis

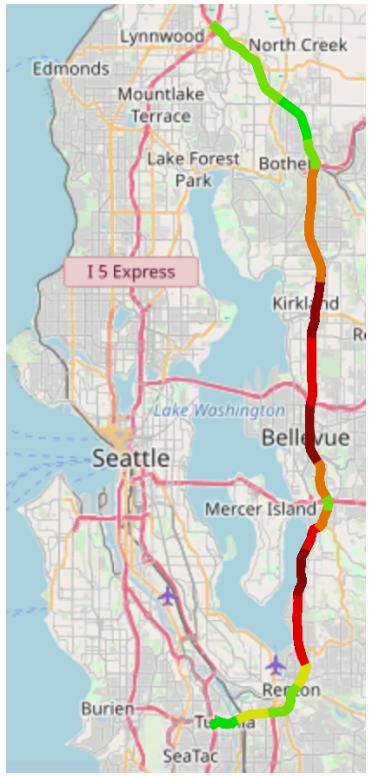
- I-405
- EB method

Visualization:

- Color
- Geometric data



Incident Frequency



Expected incident

Potential Safety Improvement Index

Next Step

Integrate data management and data visualization

Visualization in R

Introduce R Shiny

Create a data pipeline / data platform using R & Shiny and SQL Server. (Required in the Final Project)