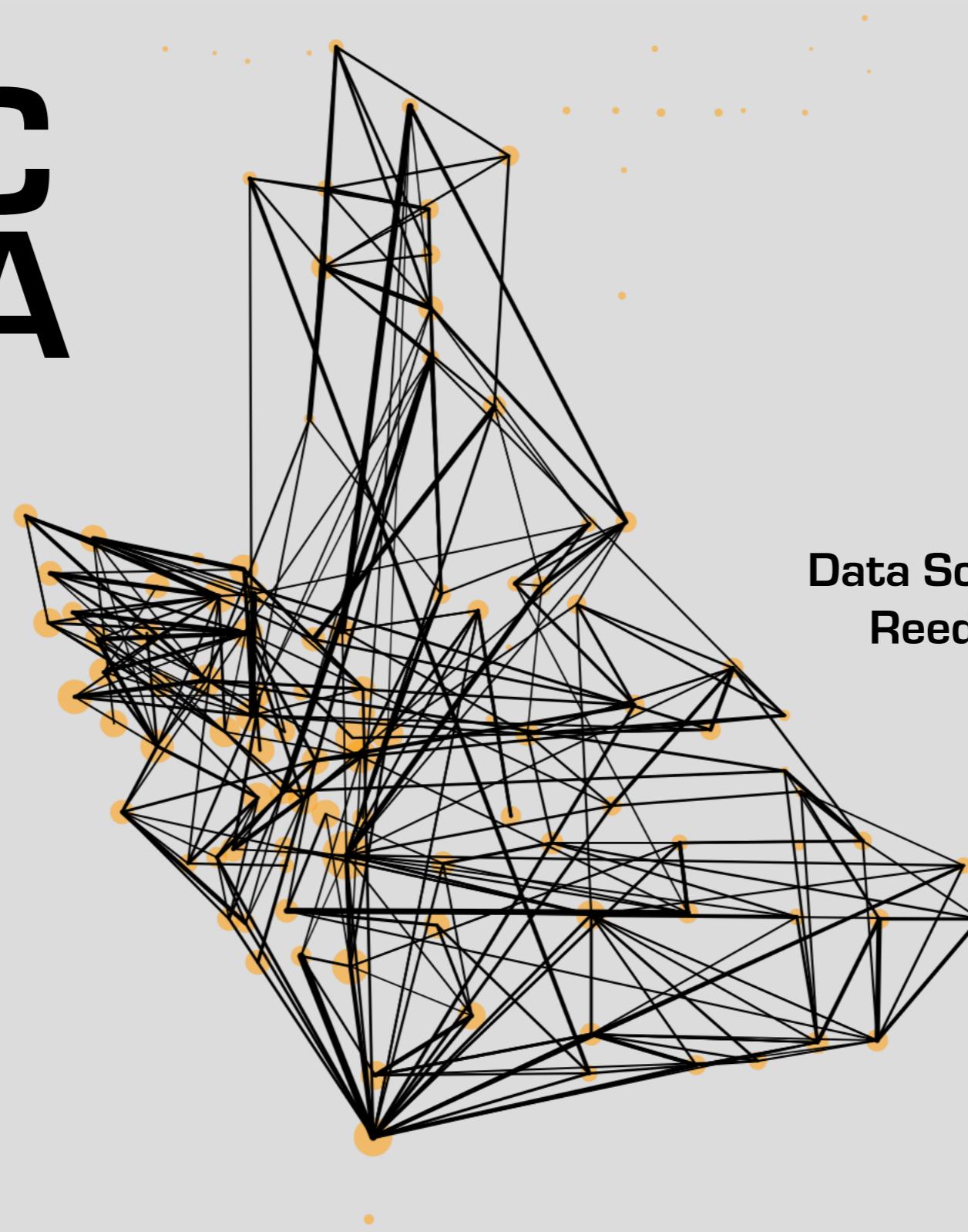


CIVIC DATA



Data Science Symposium
Reed College 5-10-18

Case Studies in Data Science: Civic Data

Spring 2018

Case Studies in Data Science: Civic Data

Spring 2018

Visualize

Case Studies in Data Science: Civic Data

Spring 2018

Visualize

Transform

Case Studies in Data Science: Civic Data

Spring 2018

Visualize

Tidy

Transform

Case Studies in Data Science: Civic Data

Spring 2018

Visualize

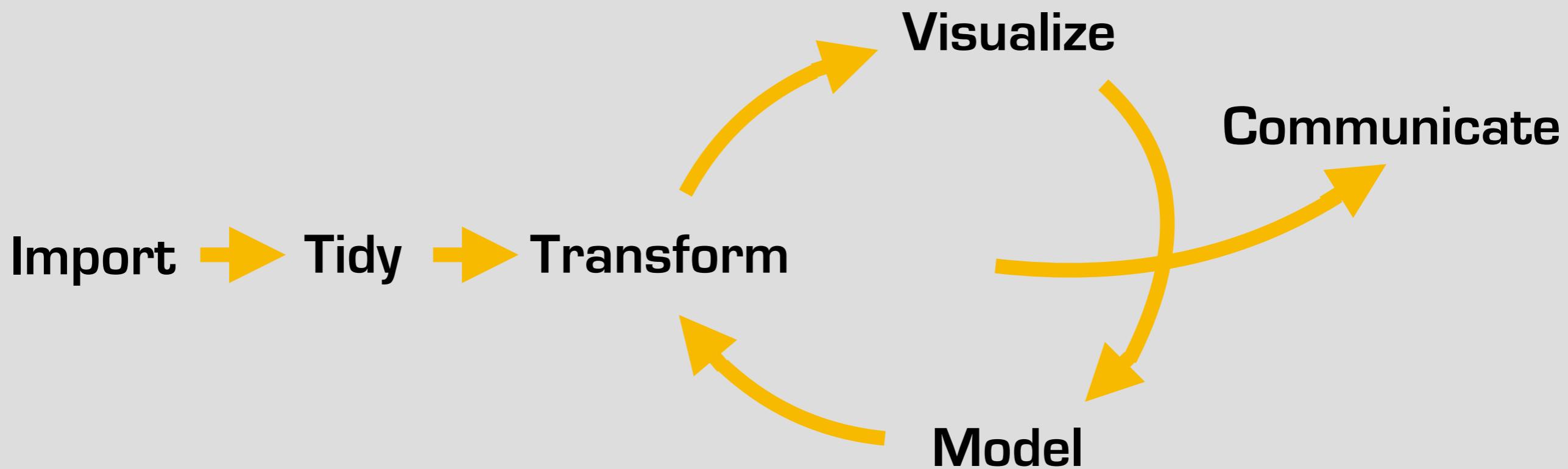
Import

Tidy

Transform

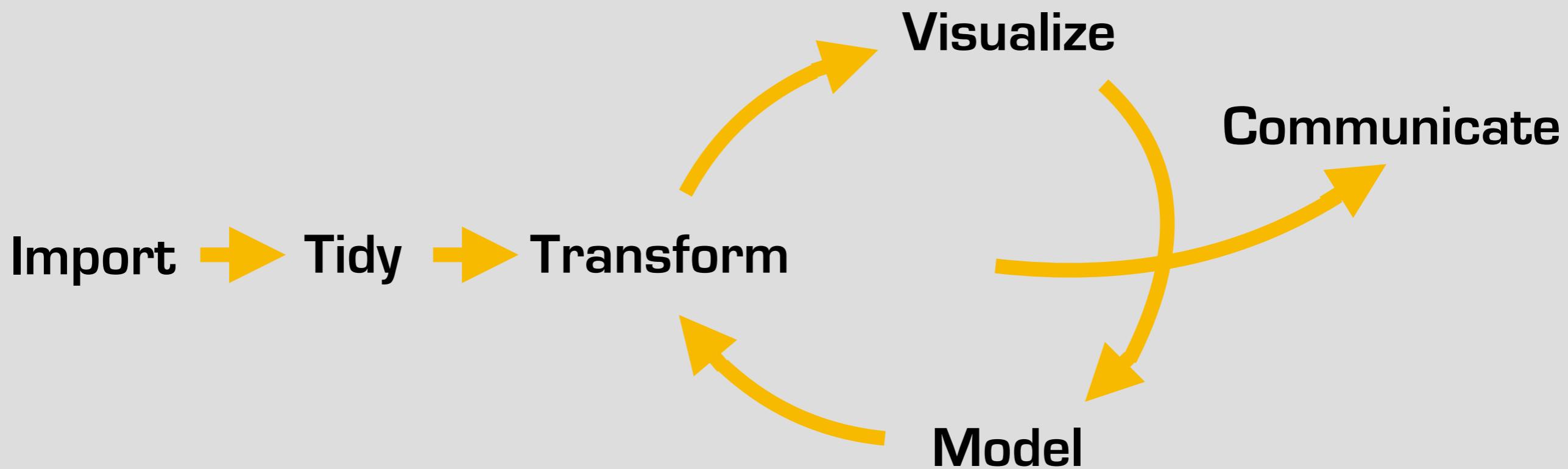
Case Studies in Data Science: Civic Data

Spring 2018



Case Studies in Data Science: Civic Data

Spring 2018



8 weeks: cycle of data science

5 weeks: group projects

Our Toolkit

- R Tidyverse
(computation)



Our Toolkit

- R Tidyverse
(computation)

- GitHub
(collaboration)

<http://github.com/ds-civic-data>



Features

Business

Explore

Marketplace

Pricing

This organization

Search



Data Science: Civic Data

A case studies course at Reed College, Portland Oregon.

Repositories 10

People 3

Projects 0

Pinned repositories

[course-materials](#)

Course materials for Math 241 at Reed College.

HTML ★1 5

Search repositories...

Type: All ▾

Language: All ▾

[news-sentiment](#)

R Updated 22 seconds ago



[sd-pdx-sea](#)

R Updated 23 minutes ago



[bikeshare-2](#)

Modeling and Visualizing Historic Portland Bike Share Ride Data

R Updated 29 minutes ago



Our Toolkit

- R Tidyverse
(computation)

- GitHub
(collaboration)

<http://github.com/ds-civic-data>

The screenshot shows the GitHub repository page for 'ds-civic-data/bikeshare-2'. The top navigation bar includes links for Features, Business, Explore, Marketplace, Pricing, and a search bar. Below the header, the repository name 'ds-civic-data / bikeshare-2' is displayed, along with tabs for Code, Issues (1), Pull requests (0), Projects (0), and Insights. A brief description of the repository states 'Modeling and Visualizing Historic Portland Bike Share Ride Data'. Key statistics shown are 65 commits, 1 branch, and 0 releases. A pull request from 'Zeki Ryan Kazan' titled 'Delete ugly stuff' is listed, along with other commits related to moving code, adding data, and merging branches. The README.md file is also visible.

ds-civic-data / bikeshare-2

Code

Issues 1

Pull requests 0

Projects 0

Insights

Modeling and Visualizing Historic Portland Bike Share Ride Data

65 commits

1 branch

0 releases

Branch: master ▾

New pull request

Zeki Ryan Kazan Delete ugly stuff

code Move to code folder

data Add temp data

misc Merge branch 'master' of https://github.com/ds-civic-data/bikeshare-2

sandbox Move to sandbox

README.md tab 1 still needs better explanation

bikeshare-2.Rproj committing to fix thing

README.md

bikeshare-2: Modeling and Visualizing Portland Bike Share Ride Data

Nathan Feldman, Zeki Kazan, Canyon Foot

We are creating a shiny app that will allow users to explore Biketown statistics. Biketown has a [complete ride data](#) for the entire history of the program. Each ride includes distance, start time, end time, and method, to which we added variables like rainfall and temperature for the hour of the ride.

Our Toolkit

- R Tidyverse
(computation)

- GitHub
(collaboration)

- Slack
(communication)

The screenshot shows a Slack interface. At the top, it says "Data Science: Ci... Andrew Bray". Below that is a "New Threads" section. Under "Channels", there is a list of channels: # activities, # bikeshare-1, # bikeshare-2, # counting-the-homeless, # emer-response-time, # general (which is highlighted in green), # nomadic-herders, # random, # rquestions, # sf-pdx-seattle, # the-big-pipe, and # urban-forest. Below this is a "Direct Messages" section with a list of users: slackbot, Andrew Bray (you), Aaron Till, Canyon Foot, Clark Chang, Frank Ga..., Emerson Webb, Frank Gaunt, Giovanni Corti, Kaelyn Rosenberg, kbott, Kyu Yeon Cho, Lirui Jiao, Lucas Williams, Nathan Feldman, Ryan Kim, and Simon Couch.

The screenshot shows a message thread in the "#general" channel. A post from Andrew Bray at 3:36 PM says: "If your group is building a shiny app, please try moving it over this evening to our shiny server (a second RStudio server that we run to just hosts apps). 1. Go to shiny.reed.edu/r and sign in. 2. You should see a folder called `ShinyApps`. Click inside there and create a new folder with the name of your app. 3. Drop all of your material inside that folder (via the upload button) sure that any paths to data set are correct. I believe the r script will app may need to be retitled to `app.r`. 4. Click over to <https://shiny.reed.edu/s/users/abray/hello/> (plug in your username for `abray` and the name of the folder that your app is in for `hello`. Once you have it working (or if you run into any problems) please me a direct message." This post has 3 replies and was last updated 1 day ago. Another post from Andrew Bray at 10:36 AM says: "Hi andrew, my groups draft of our white paper is uploaded to our github repository, just thought Id let you know." A post from Josephine Baker at 12:39 PM says: "Practice presentations, both this afternoon and tomorrow morning be in Library 203, one floor beneath my office." A post from kbott at 5:03 PM says: "I can't make it to presentations tomorrow, but I hope they go well congrats on doing great work this term. See you next fall (okay, no seniors)."

Our Toolkit

- R Tidyverse
(computation)

- GitHub
(collaboration)

- Slack
(communication)

Data Science: Ci... Andrew Bray

New Threads

Channels

- # activities
- # bikeshare-1
- # bikeshare-2
- # counting-the-homeless
- # emer-response-time
- # general
- # nomadic-herders
- # random
- # rquestions
- # sf-pdx-seattle
- # the-big-pipe
- # urban-forest

Direct Messages

- slackbot
- Andrew Bray (you)
- Aaron Till
- Canyon Foot
- Clark Chang, Frank Gaunt
- Emerson Webb
- Frank Gaunt
- Giovanni Corti
- Kaelyn Rosenberg
- kbott
- Kyu Yeon Cho
- Lirui Jiao
- Lucas Williams
- Nathan Feldman
- Ryan Kim
- Simon Couch

#general

☆ | 8 29 | 0 | Company-wide, and add to emoji Monday, May 7th conferencing capabilities).

Reminder: draft deliverables (paper/poster/app) are due Wednesday morning.

Tuesday, May 8th

Andrew Bray 3:36 PM If your group is building a shiny app, please try moving it over this RStudio server that we run to ...

nyApps. Click inside there and f your app.

at folder (via the upload button) correct. I believe the r script wi ...

.edu/s/users/abray/hello/ (and the name of the folder that you ...

un into any problems) please ...

All-Time Usage

Messages Sent

4,560

out of 10,000 messages

↑ 2,049 messages (+81.6%) over the last 30 days.

Josephine Baker 10:36 AM Hi andrew, my groups draft of our white paper is uploaded to our github repository, just thought Id let you know.

Andrew Bray 12:39 PM Practice presentations, both this afternoon and tomorrow morning, be in Library 203, one floor beneath my office.

kbott 5:03 PM I can't make it to presentations tomorrow, but I hope they go well congrats on doing great work this term. See you next fall (okay, no seniors).

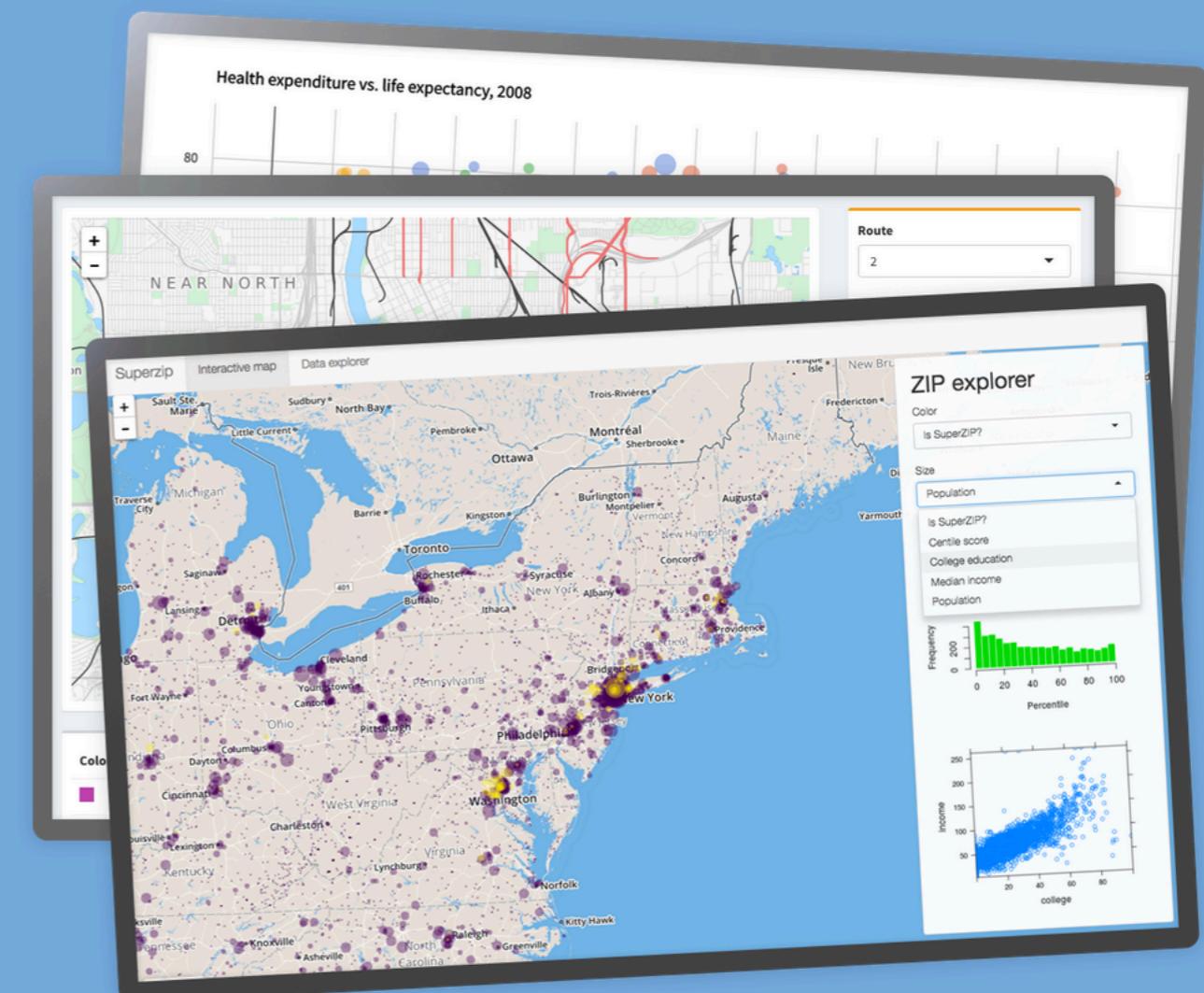
+ Message #general

Our Toolkit

- **R Tidyverse**
(computation)
- **GitHub**
(collaboration)
- **Slack**
(communication)
- **Shiny**
(extramural communication)

Interact. Analyze. Communicate.

Take a fresh, interactive approach to telling your data story with Shiny. Let users interact with your data and your analysis. And do it all with R.



**CIVIC
DATA**

Why **CIVIC
DATA**?

Why CIVIC
DATA?

Transparency &
Accountability

Why CIVIC
DATA?

Transparency &
Accountability

Civic
Service

Why CIVIC DATA?

Transparency &
Accountability

Civic
Service

Civic
Engagement

Projects

- 1. PDX Tree Planting**
- 2. Urban Forest**
- 3. News Sentiment**
- 4. Migration, Occupation, and Real Estate**
- 5. National and Regional Trends in Homelessness**
- 6. gbfs: an R package for bikeshare data**
- 7. Biketown: modeling Portland's bikeshare network**
- 8. Nomadic Herders and Weather Risk**

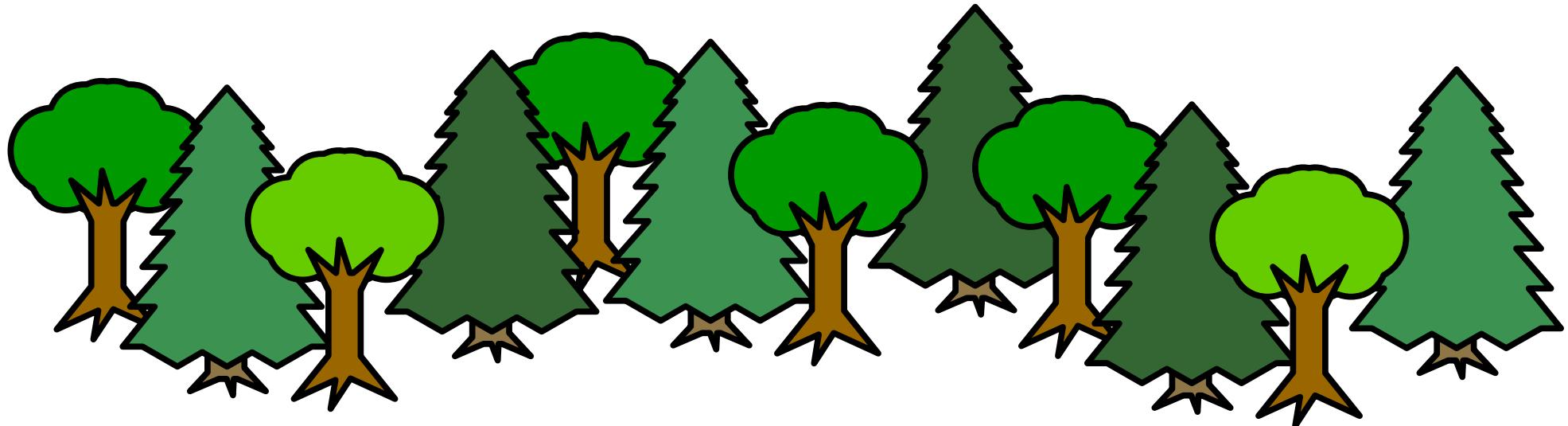
CascadiaRConf

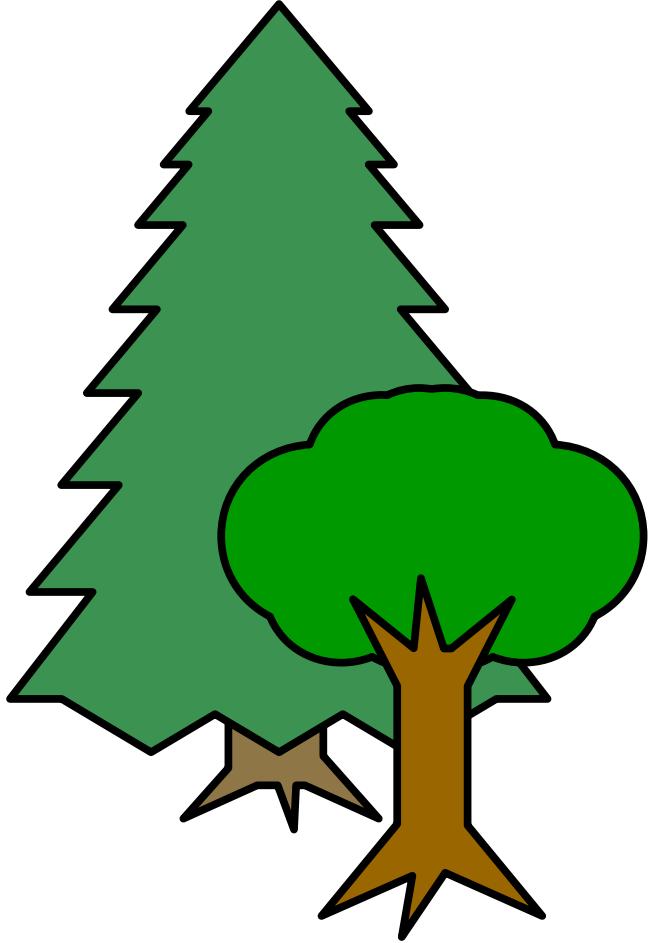


Cascadia
R Conf

June 2, 2018
Portland, OR

PDX Tree Planting

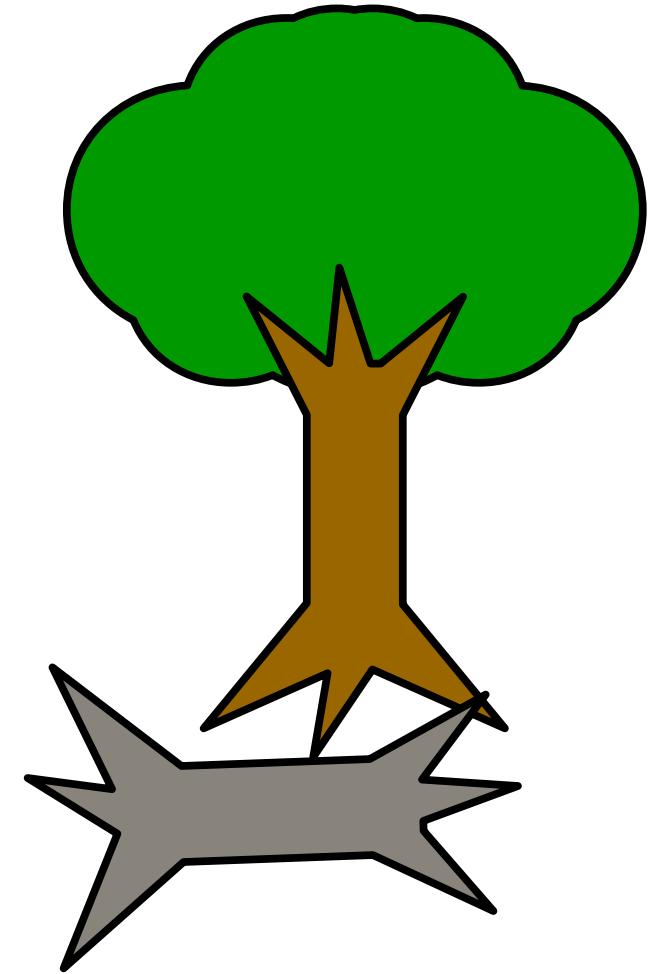




**Are we planting
the right types
of trees?**



**Are planting
efforts
equitable?**



**Which trees
persist?**

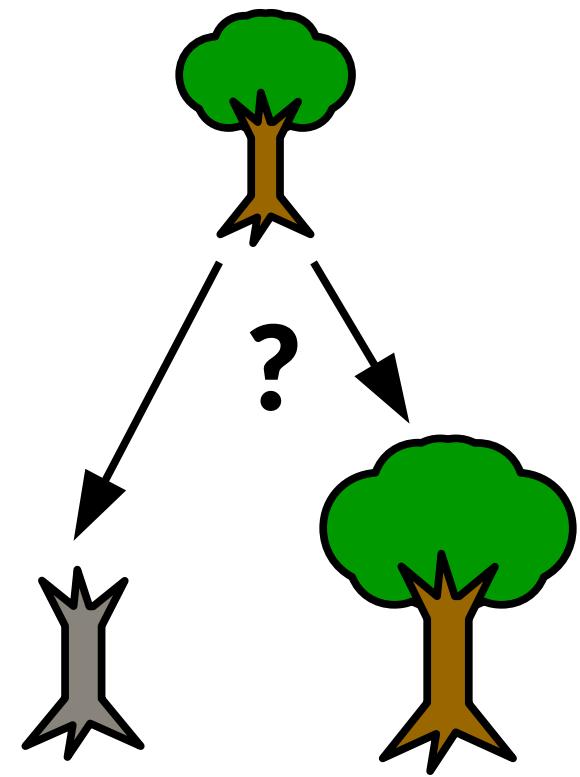
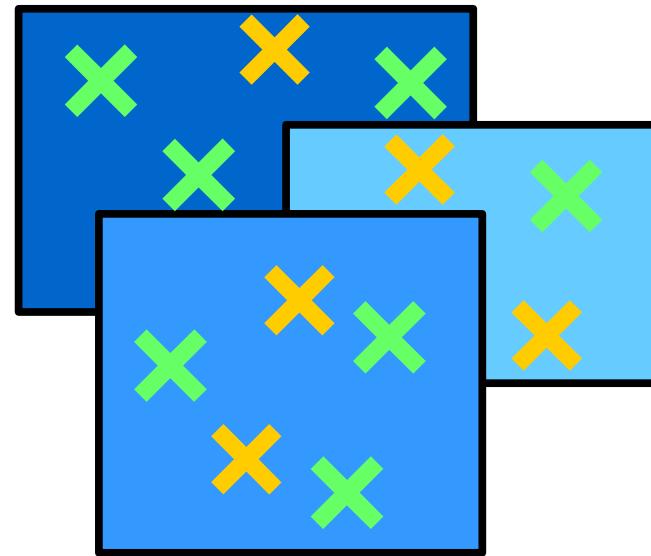
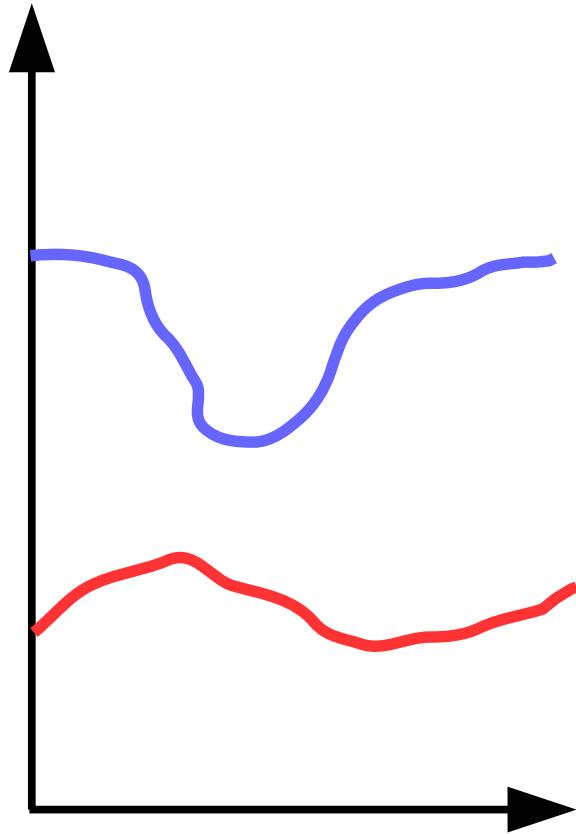
planting data

income data

Name	Year	Size	Native	Type	FIPS	Income	Persist
Norway maple	1999	M	False	Decid.	4105...	Med-High	True
Douglas fir	2005	L	True	Ever.	4105...	Low	False

census tracts

tree survey



**Assessing tree
quality trends
over time**

**Mapping trees
against income**

**Creating a
model to
predict
persistence**



Quality has improved, but needs work

Planting has been fairly equitable

Large, deciduous trees persist the most

URBAN FOREST

URBAN-FOREST

Q: WHAT IS THE RELATIONSHIP BETWEEN
DEMOGRAPHIC VARIABLES AND TREES?

URBAN-FOREST

Q: WHAT IS THE RELATIONSHIP BETWEEN
DEMOGRAPHIC VARIABLES AND TREES?



ACS CENSUS
DATA

URBAN-FOREST

Q: WHAT IS THE RELATIONSHIP BETWEEN
DEMOGRAPHIC VARIABLES AND TREES?



ACS CENSUS
DATA



CANOPY
COVERAGE

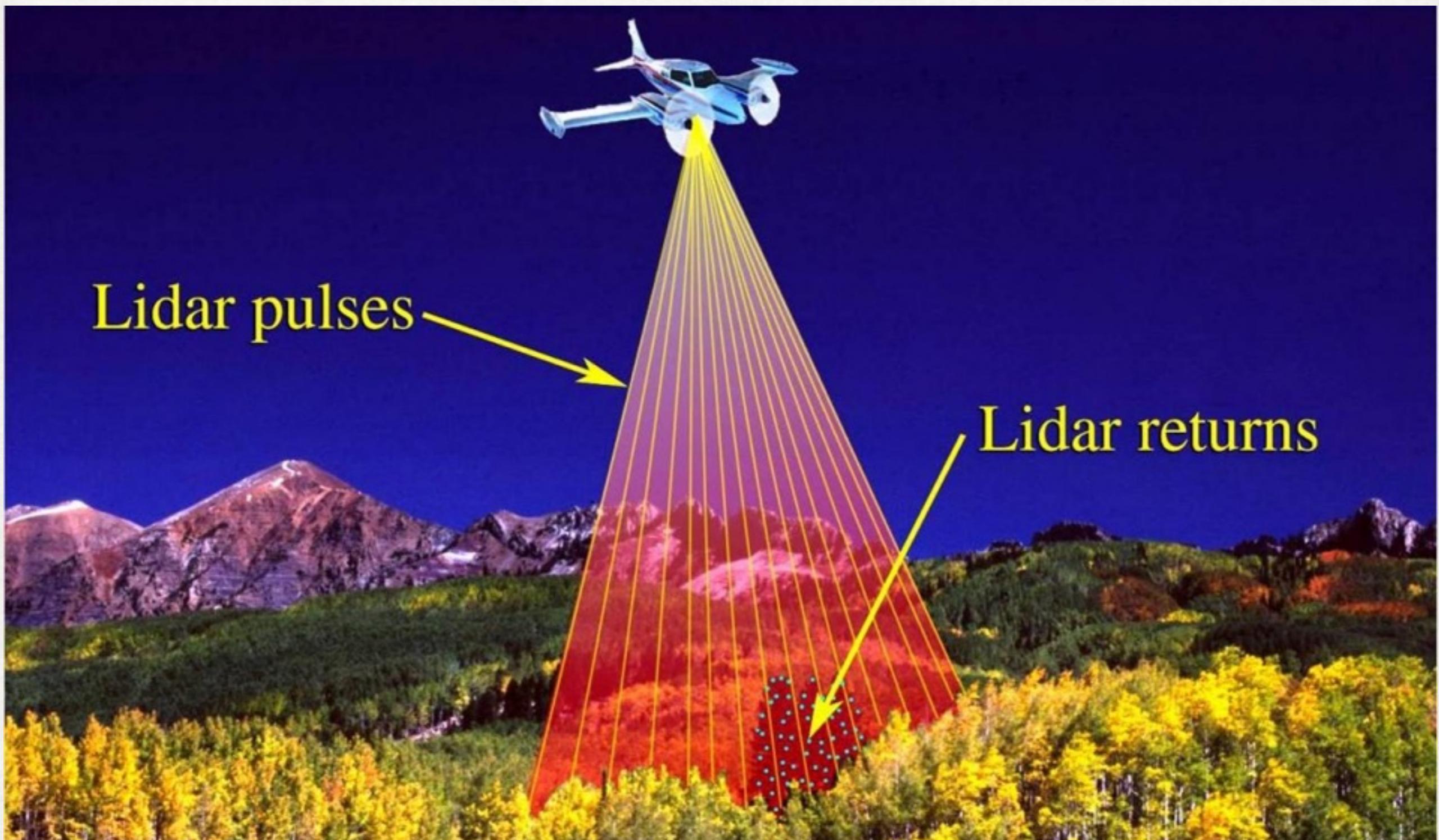
URBAN-FOREST

Q: WHAT IS THE RELATIONSHIP BETWEEN
DEMOGRAPHIC VARIABLES AND TREES?

ACS CENSUS
DATA

CANOPY
COVERAGE

LIDAR DATA





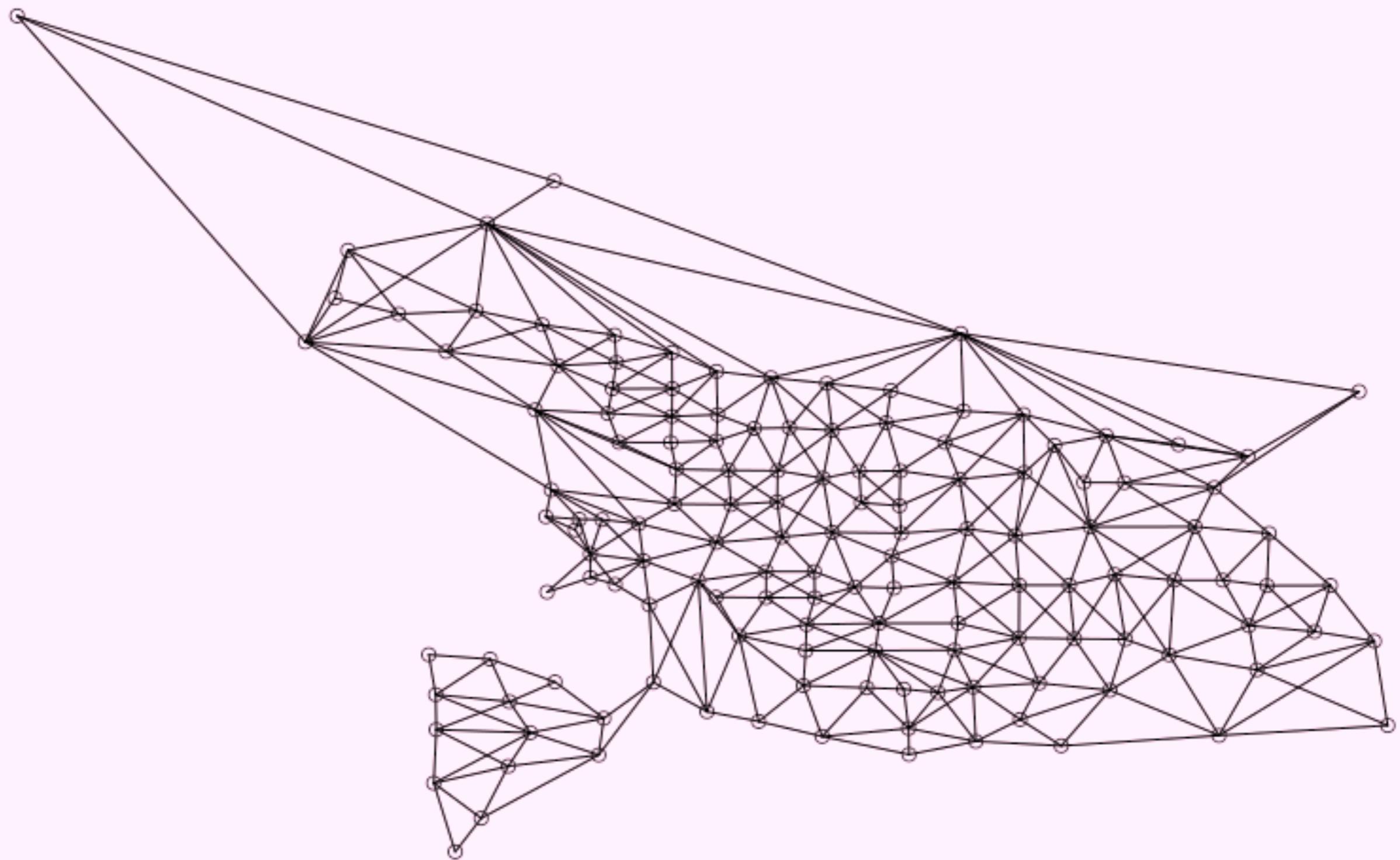
URBAN-FOREST

**Q: WHAT IS THE RELATIONSHIP BETWEEN
DEMOGRAPHIC VARIABLES AND TREES?**

URBAN-FOREST

Q: WHAT IS THE RELATIONSHIP BETWEEN
DEMOGRAPHIC VARIABLES AND TREES?







News Sentiment Analysis

Emerson Webb, Richard Pham, and Henry Knopf

with special thanks to NewsAPI.org

Purpose

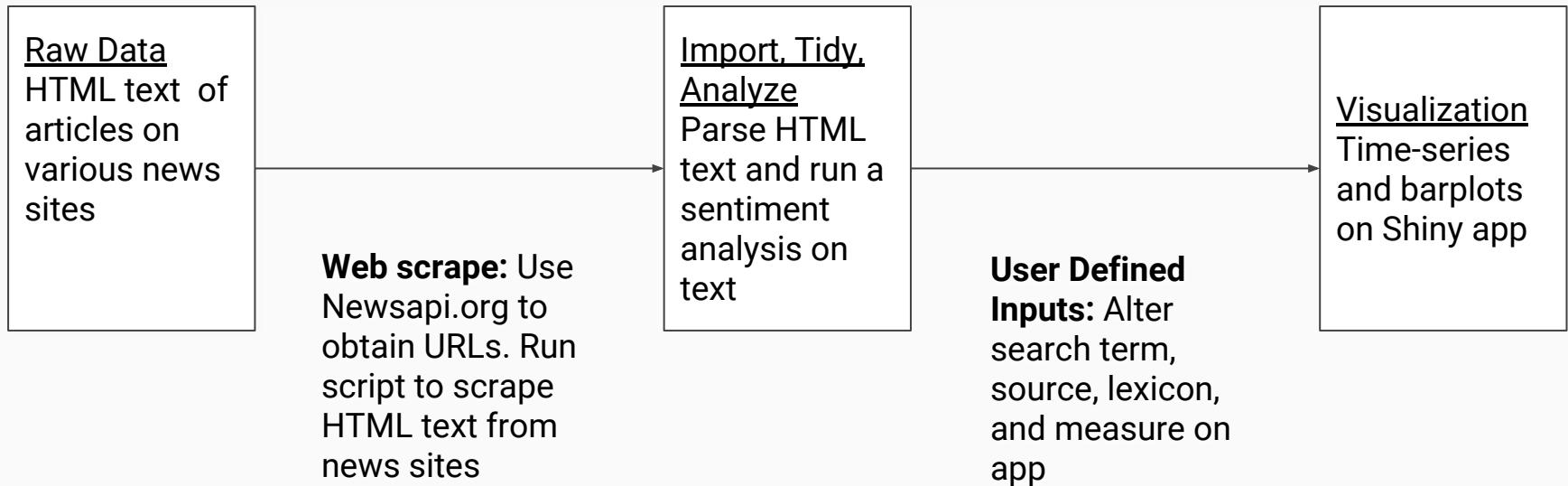
- ❖ Visualize and compare sentiments of news articles on different news sites.
- ❖ Observe if a partisan trend exists across news sites
- ❖ Future Goal:
 - Automate script to collect long-term data and integrate with Shiny app
 - Make sentiment-scoring more precise

Raw Data

The screenshot shows a news article from CBS News. The header includes the CBS News logo, navigation links for NEWS, SHOWS, VIDEO, CBSN, MORE, and a search icon. The main title of the article is "Report: Gina Haspel offered to withdraw nomination to head CIA before confirmation hearing". Below the title are sharing options: Share, Tweet, Reddit, Flipboard, and Email. A timestamp indicates the article was last updated on May 7, 2018, at 7:19 AM EDT. The text of the article discusses President Trump's pick to lead the CIA, Gina Haspel, who reportedly offered to withdraw her nomination after a White House meeting focusing on her past role in interrogation programs. It cites The Washington Post and four senior U.S. officials. The text also mentions that Haspel had advised the White House she would withdraw to avoid a potentially harsh confirmation hearing. A sidebar notes that officials told the Post she didn't want to risk damaging the reputation of the CIA and herself in the process. Another official referred to Haspel as the "next Ronny Jackson," referring to Mr. Trump's failed nominee for Veterans Affairs secretary. The final paragraph states that CBS News senior investigative producer Pat Milton reports that a source close to the matter said that Haspel's nomination "is going forward." An administration official also told CBS News that her nomination remains on track.

An example of a URL gathered. The text gathering function only captures the main body of text.

Data Pipeline



Shiny App

News Website Sentiment Analysis

Search Term

News Sources

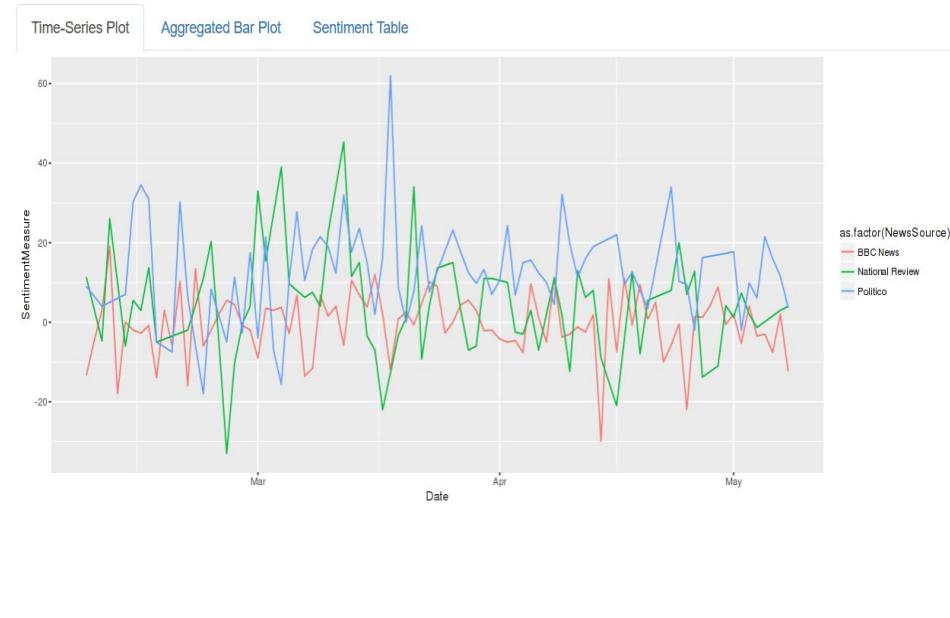
the-guardian-uk national-review the-huffington-post bbc-news
politico

Sentiment Lexicon

Score Type

Sentiment Measure

refresh



Migration, Occupation, and Real Estate in Portland

By: Josephine Baker, Lucas Williams, Kyu Yeon Cho

General Questions

- How has out-of-state migration to Portland changed over time?
- Specifically, what do patterns of migration from the Bay Area and Puget Sound Area look like?
- How has economic disparity changed in Portland the last ten years?
- How have housing prices changed over the last ten years?
- How might patterns of migration, economic distribution, and housing prices relate?

Real Estate Data
Source: Portland Metro

Migration Data
Source: IRS

Job/ Economic Data
Source: Census Bureau

Unit of observation:
A single property

Variables: Zipcode,
County, Sale date, Sale
Price, Land Use

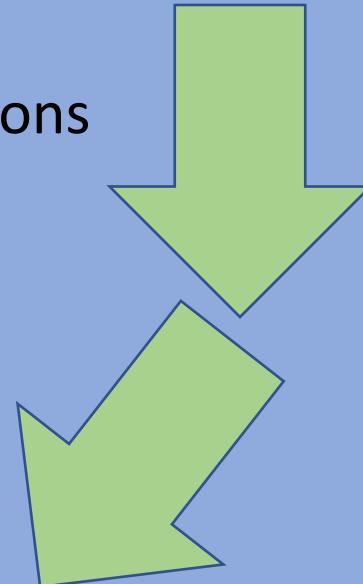
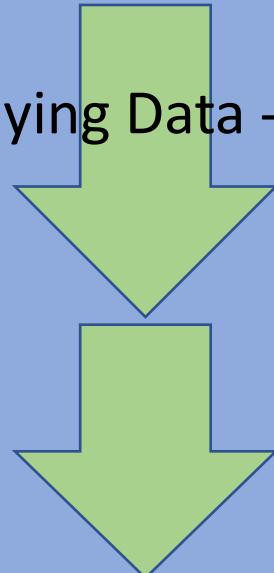
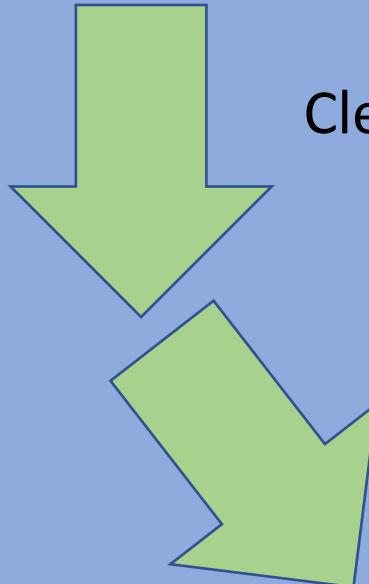
Unit of Observation:
County to County

Variables: Origin
County, Destination
County, Number, AGI

Unit of Observation:
County

Variables: Gini
Coefficient, Age, Job
Type and Number

Cleaning and Tidying Data ---> Visualizations



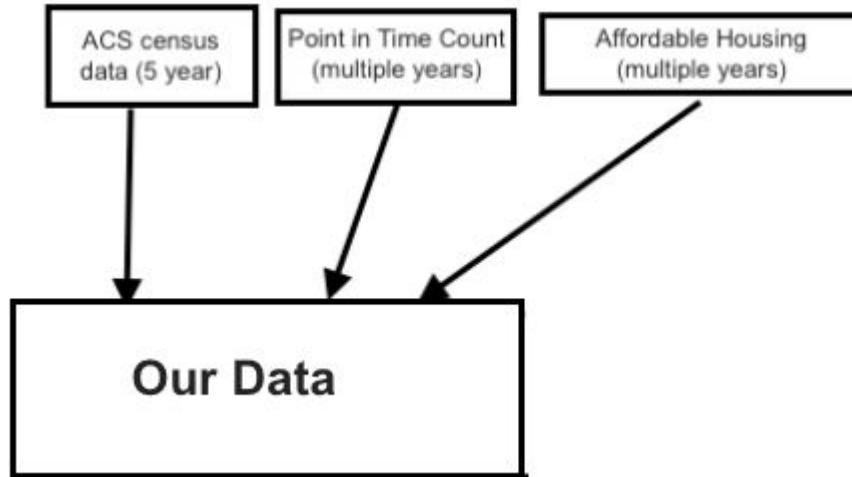
MODELLING

Regional and National Trends in Homelessness

Questions

- What are the regional (Oregon) and national trends of Homelessness?
- Which factors are best at explaining Homelessness in the aggregated level?

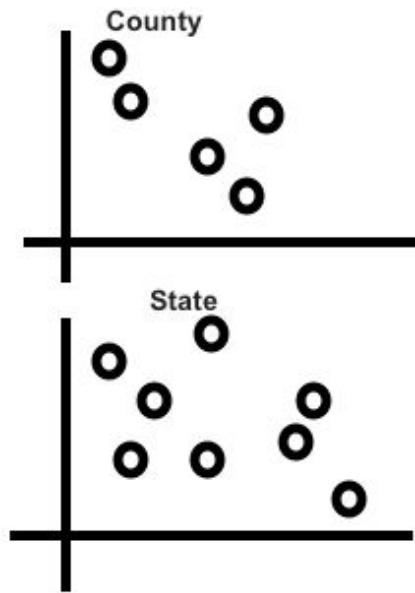
Our Data



What our Shiny offers

Options:

- Homelessness Category
- X Variables for OR counties
- X Variables for States
- State Selection
- Fit Line
- X Variable for AV plot



gbfs: R Package

General Bikeshare Feed Specification

Kaelyn M. Rosenberg & Simon P. Couch

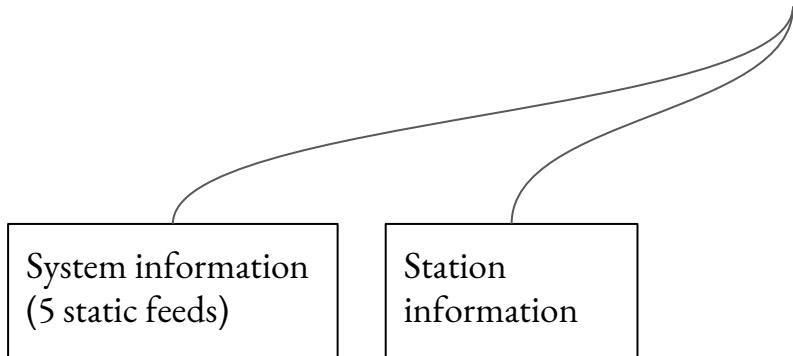
What is gbfs?

What is gbfs?

```
{"last_updated":1525975173,"ttl":180,"data":{"stations": [{"station_id":"hub_1512","num_bikes_available":17,"num_bikes_disabled":0,"num_docks_available":1,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1513","num_bikes_available":8,"num_bikes_disabled":0,"num_docks_available":10,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1514","num_bikes_available":15,"num_bikes_disabled":0,"num_docks_available":3,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1515","num_bikes_available":4,"num_bikes_disabled":0,"num_docks_available":13,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1516","num_bikes_available":8,"num_bikes_disabled":0,"num_docks_available":8,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1517","num_bikes_available":5,"num_bikes_disabled":0,"num_docks_available":9,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1518","num_bikes_available":9,"num_bikes_disabled":0,"num_docks_available":7,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1519","num_bikes_available":11,"num_bikes_disabled":0,"num_docks_available":7,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1521","num_bikes_available":5,"num_bikes_disabled":0,"num_docks_available":13,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1522","num_bikes_available":2,"num_bikes_disabled":0,"num_docks_available":16,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1523","num_bikes_available":11,"num_bikes_disabled":0,"num_docks_available":7,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1524","num_bikes_available":4,"num_bikes_disabled":0,"num_docks_available":8,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1527","num_bikes_available":0,"num_bikes_disabled":0,"num_docks_available":10,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1528","num_bikes_available":6,"num_bikes_disabled":0,"num_docks_available":12,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1529","num_bikes_available":0,"num_bikes_disabled":0,"num_docks_available":18,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1535","num_bikes_available":7,"num_bikes_disabled":0,"num_docks_available":11,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}]}]
```

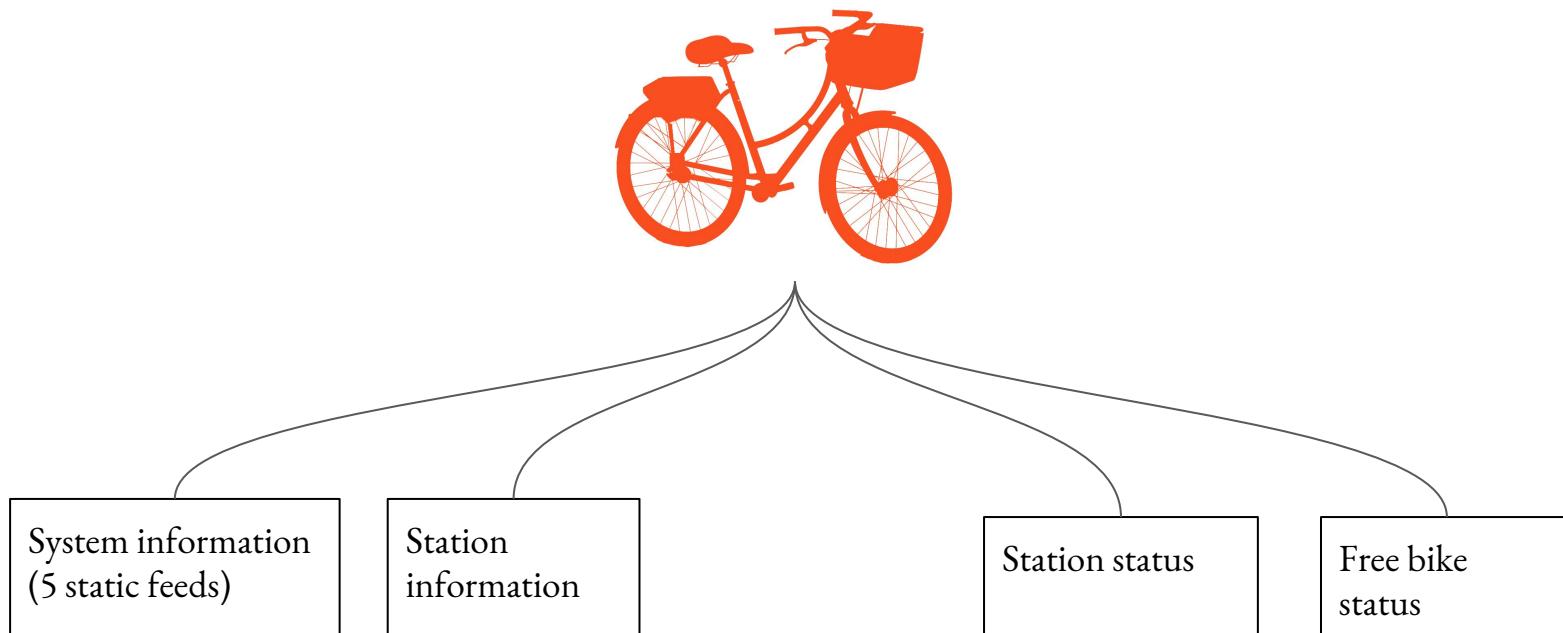
What is gbfs?

```
{"last_updated":1525975173,"ttl":180,"data":{"stations": [{"station_id":"hub_1512","num_bikes_available":17,"num_bikes_disabled":0,"num_docks_available":1,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1513","num_bikes_available":8,"num_bikes_disabled":0,"num_docks_available":10,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173},
```



What is gbfs?

```
{"last_updated":1525975173,"ttl":180,"data":{"stations": [{"station_id":"hub_1512","num_bikes_available":17,"num_bikes_disabled":0,"num_docks_available":1,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173}, {"station_id":"hub_1513","num_bikes_available":8,"num_bikes_disabled":0,"num_docks_available":10,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1525975173},
```



```
get_gbfs(city = "portland", directory = "pdx")
```

```
get_gbfs(city = "portland", directory = "pdx")
```

- Makes a folder called “pdx”
- Finds biketown’s URL
- Finds which feeds biketown supplies
- Downloads all feeds
 - Converts .json to .rds
 - Specifies column classes
 - Mutates new date & time columns
 - Saves with informative file names

`get_gbfs(city = "portland", directory = "pdx")`

- Makes a folder called “pdx”
- Finds biketown’s URL
- Finds which feeds biketown supplies
- Downloads all feeds
 - Converts .json to .rds
 - Specifies column classes
 - Mutates new date & time columns
 - Saves with informative file names

hub	# bikes available	time	...
1512	4	8:00 a.m.	...
1513	7	8:00 a.m.	...
...

`get_gbfs(city = "portland", directory = "pdx")`

- Makes a folder called “pdx”
- Finds biketown’s URL
- Finds which feeds biketown supplies
- Downloads all feeds
 - Converts .json to .rds
 - Specifies column classes
 - Mutates new date & time columns
 - Saves with informative file names

hub	# bikes available	time	...
1512	4	8:00 a.m.	...
1513	7	8:00 a.m.	...
...
1512	3	8:05 a.m.	...
1513	4	8:05 a.m.	...
...

Additions to the Data

X1	StartDate	PRCP	SNWD	TAVG	TMAX	TMIN	RouteID	PaymentPlan	StartHub
13	2016-07-19	0	0	68	78	62	1282179	Subscriber	NW 8th at Everett
14	2016-07-19	0	0	68	78	62	1282182	Subscriber	NW 2nd at Everett
15	2016-07-19	0	0	68	78	62	1282200	Subscriber	SW 3rd at Ankeny
16	2016-07-19	0	0	68	78	62	1282204	Subscriber	SW Stark at Burnside
17	2016-07-19	0	0	68	78	62	1282205	Subscriber	NA
18	2016-07-19	0	0	68	78	62	1282208	Subscriber	SW 3rd at Ankeny
19	2016-07-19	0	0	68	78	62	1282218	Subscriber	N Williams at Fremont
20	2016-07-19	0	0	68	78	62	1282243	Subscriber	SW 5th at Main

...

StartLatitude	StartLongitude	StartTime	EndHub	EndLatitude	EndLongitude	EndDate	EndTime
---------------	----------------	-----------	--------	-------------	--------------	---------	---------

...

BikeID	BikeName	Distance_Miles	Duration	RentalAccessPath	MultipleRental	hour	daily_total	rainfall
--------	----------	----------------	----------	------------------	----------------	------	-------------	----------

Linear Regression Output

Variable Selection

Select Variable to Predict

Average Duration

Select Variable for Model

Rainfall

DayOfWeek

Season

Temperature

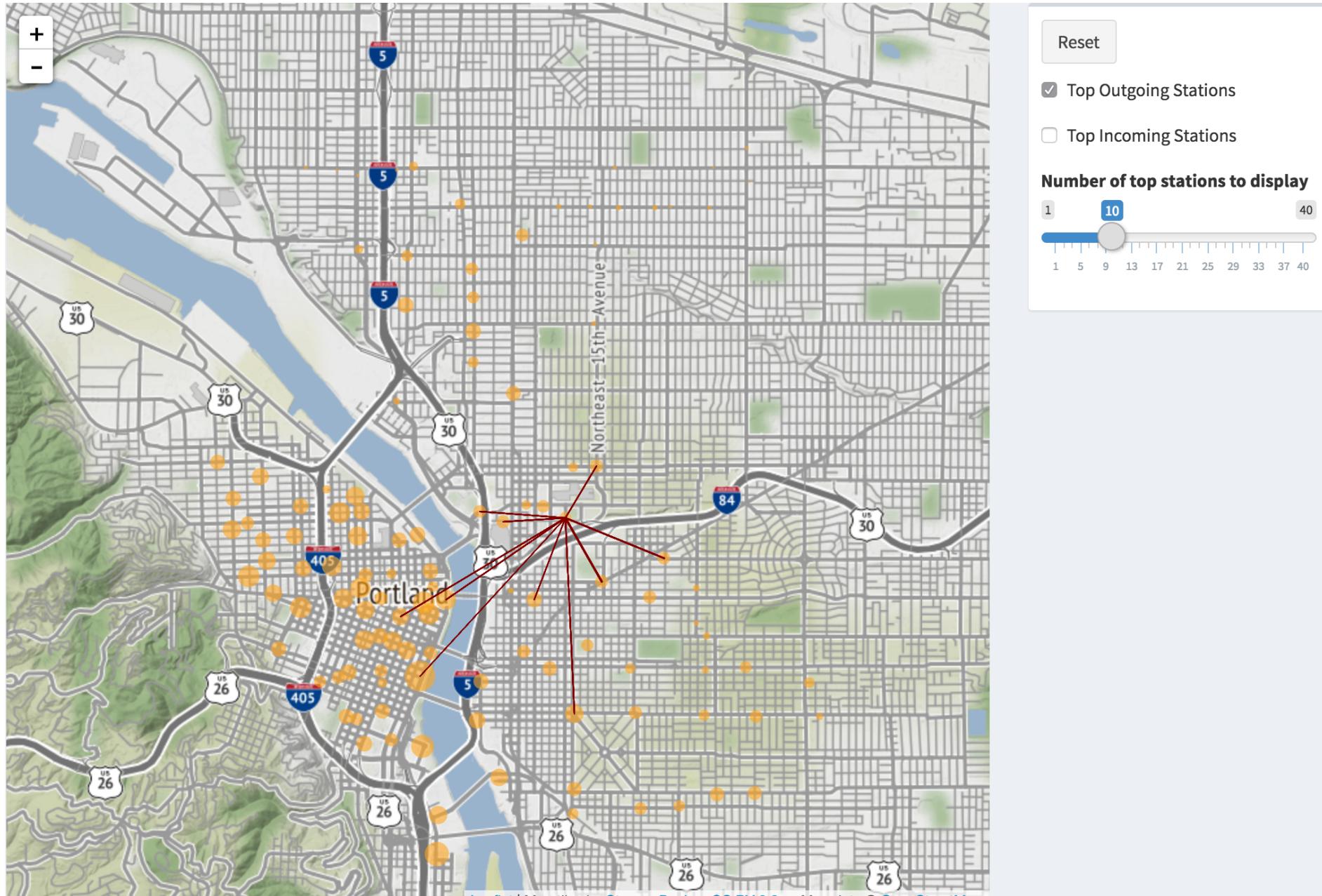
Output

Predictor	Effect	Significance
Rainfall	⬇️ Negative	Very Significant
DayOfWeekSunday	⬆️ Positive	Very Significant
DayOfWeekMonday	⬆️ Positive	Significant
DayOfWeekTuesday	⬇️ Negative	Not Significant
DayOfWeekThursday	⬆️ Positive	Not Significant
DayOfWeekFriday	⬆️ Positive	Very Significant
DayOfWeekSaturday	⬆️ Positive	Very Significant
SeasonSpring	⬆️ Positive	Very Significant
SeasonSummer	⬆️ Positive	Very Significant
SeasonWinter	⬆️ Positive	Not Significant
Temperature	⬆️ Positive	Very Significant

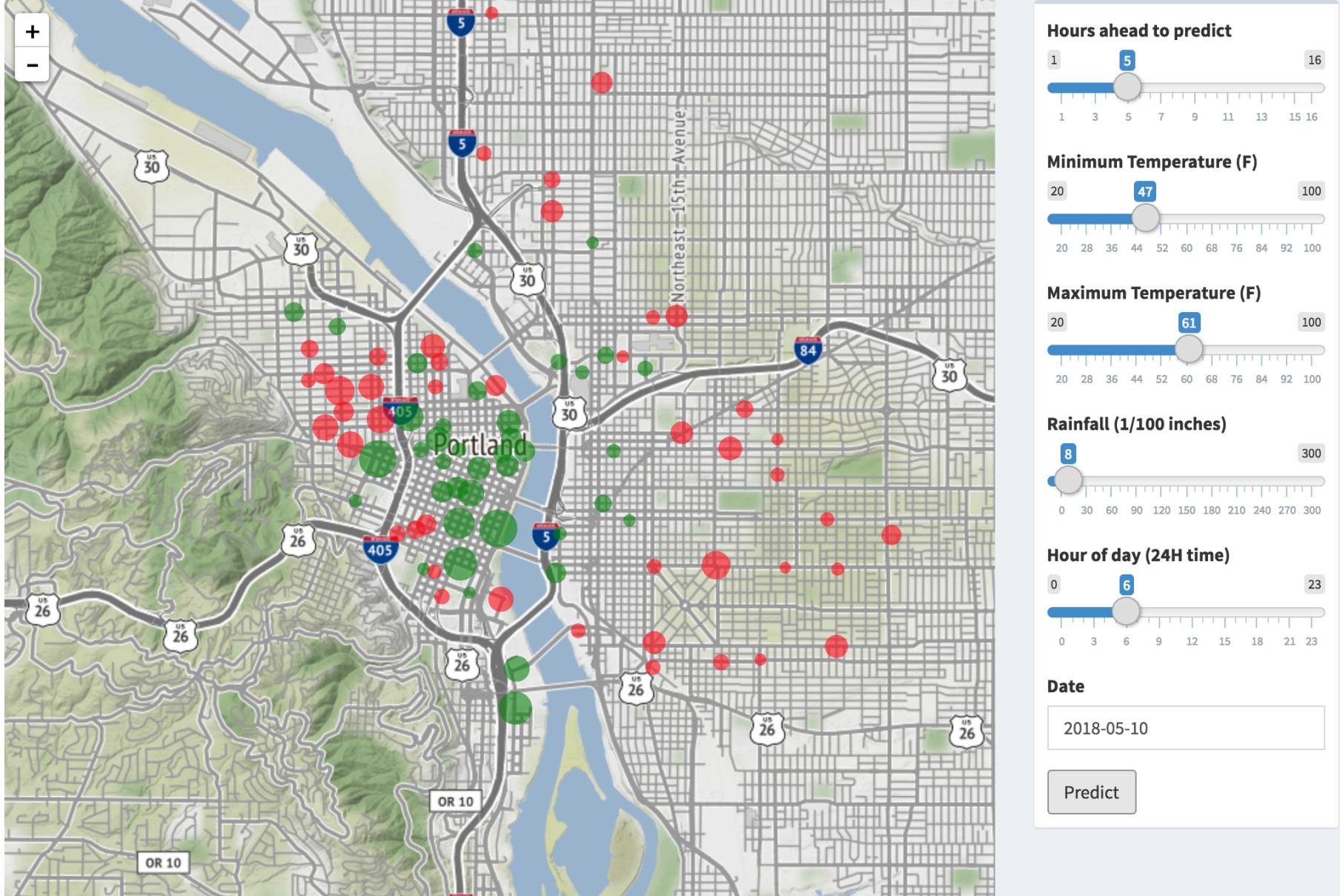
Linear Regression

Variable Selection	Prediction
Select Variable to Predict <input type="text" value="Number of Trips"/>	1006.932
Select Day of the Week <input type="text" value="Thursday"/>	
Select Level of Rainfall (.01 inch) <input type="text" value="8"/>	
Select Temperature (Degrees Fahrenheit) <input type="text" value="61"/>	
Select Season: <input type="text" value="Spring"/>	

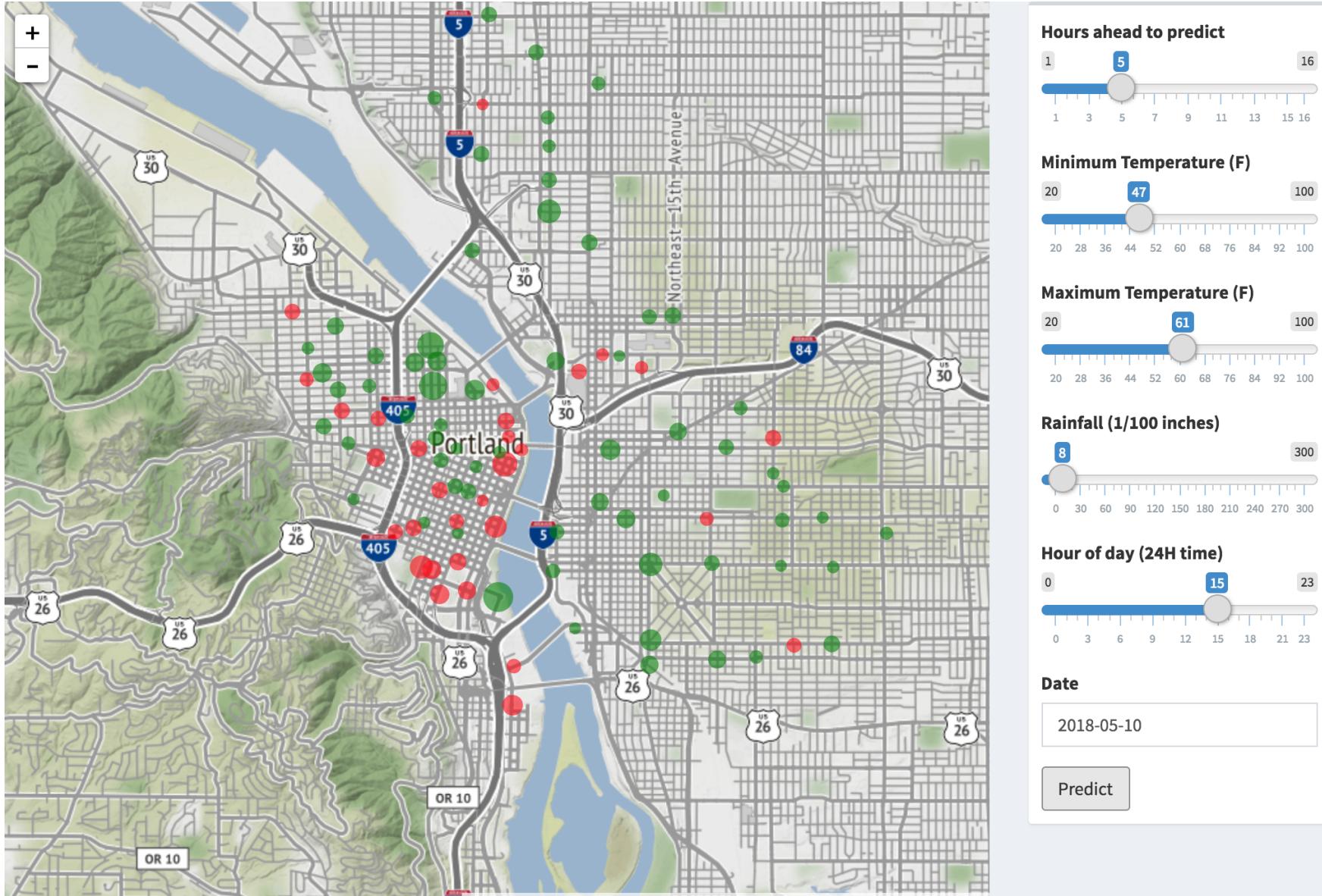
Predicting Where Bikes Travel



Changes in Station Use (6:00-11:00 AM)



Changes in Station Use (3:00-8:00 PM)



LTS-2 System

Exploratory Analysis of Service Usage Over Time and Space





(123)-555-5555

**Area code and
requested forecast**



**Weather or
pasture forecast**



LTS Service





Area code and requested forecast

Weather or pasture forecast



ID	Date	Time	Type	Message
14278	2017-04-01	08:11:21	In	62267 2
14278	2017-04-01	08:12:17	Out	4sar4:uulerheg



Areas of Focus

- Individual Users
- General Trends in System Usage Over Time
- Spatial Distribution of Usage



