Climate opinion maps

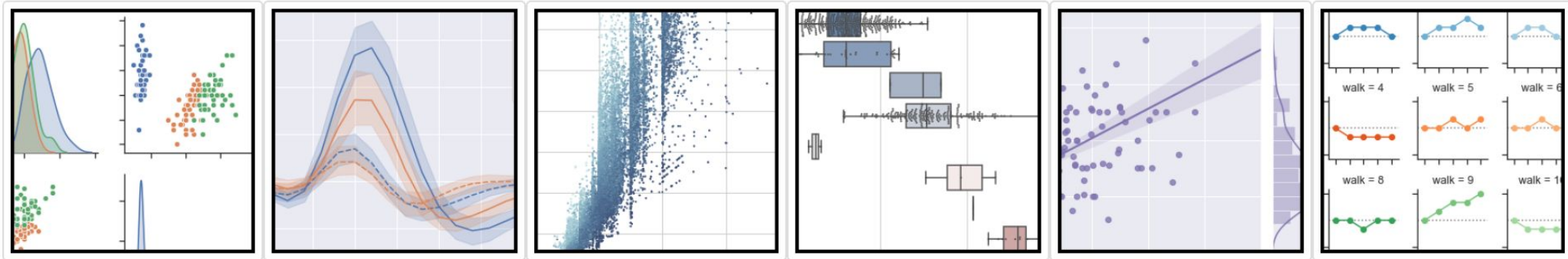**Objective:** Learning more about beliefs about climate change
- whether these might be predictors of certain types of behaviors in the United States
- whether these beliefs are linked to certain demographic variables.

This project will **create a data set** that allows researchers to **explore whether a relationship exists between beliefs about climate change, demographic characteristics and transportation behaviors** (for example walking, cycling, driving behaviors, use of transit types and number of cars owned) in the United States.

Main technology needs: Data manipulation, statistics, data visualization.

Background	Application: Need statistics functions to get statistics parameters: correlation coef/regression, and/or significance from selected variables	Drawbacks
SciPy contains modules for linear algebra, optimization, integration, and statistics. The main functionality of SciPy library is built upon NumPy, and its arrays thus make substantial use of NumPy. Provides efficient numerical routines as numerical integration, optimization, and many others via its specific submodules. The functions in all submodules of SciPy are well documented	**Stats.linregress( )** a highly specialized linear regression function **It is one of the fastest method when it comes to simple linear regression.** Apart from the fitted coefficient and intercept term, it also returns basic statistics such as R^2 coefficient and standard error.	It is fairly restricted in its flexibility as it is optimized to calculate a linear least-squares regression for two sets of measurements only. R^2
	Stats.pearsonr( ) Calculate a Pearson correlation coefficient and the p-value for testing non-correlation. This will provide us significance	Possible **bug with zero variance input** #3728 in scipy github

seaborn: statistical data visualization

Background	Application	Drawbacks
Seaborn is mostly focused on the visualization of statistical models; such visualizations include heat maps, those that summarize the data but still depict the overall distributions. Seaborn is based on Matplotlib and highly dependent on that. There is a rich gallery of visualizations including some complex types like time series, jointplots, and violin diagrams. The seaborn updates mostly cover bug fixes.	**sns.regplot( )** Plot data and a linear regression model fit. **sns.boxplot( )** Draw a box plot to show distributions with respect to categories	Functions only provides maps but no coefficients (e.g., R^2) Possible solution: **Use statsmodels.api**
	sns.heatmap( ) Plot rectangular data as a color-encoded matrix.	

Heatmap of correlations obtained from **climate opinion vs. census** variables based on counties
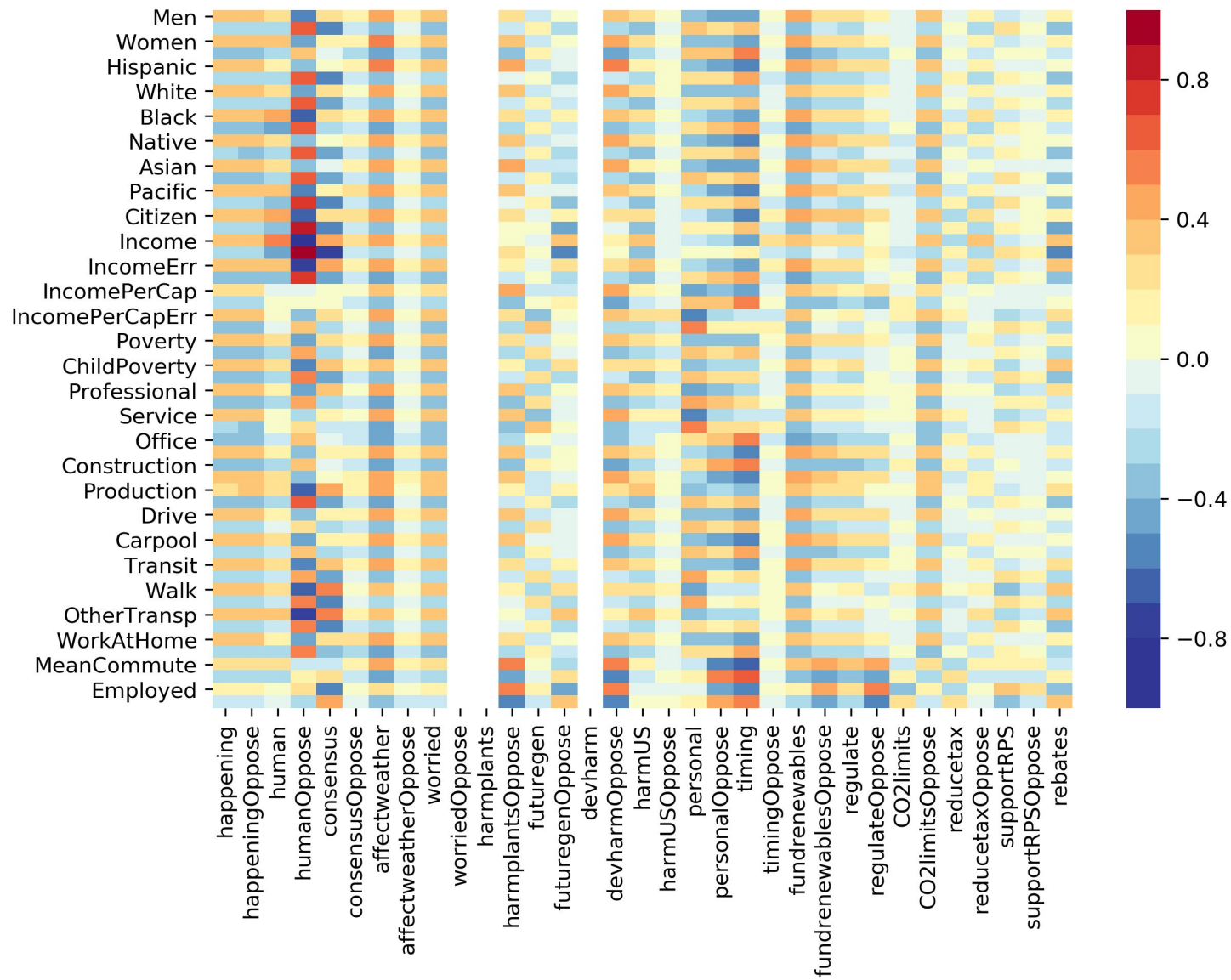
x axis: Climate opinion variables

y axis: Census variables

Altair

Technology need:
- We are looking for tools that have the option to create interactive data visualizations.

Appeal
- Declarative statistical visualization
- Beautiful plots, minimal and concise code
- Easy to make interactive plots
- Some core developers are at UW!

How to use:
Still figuring this out but syntax is very simple and there are many examples/notebooks/tutorials online.

Drawbacks:
Recently developed

Background:

Need an **interactive interface for visualizing** which allows the user to select between different variables and explore their statistical relationship.

Bokeh:

- Is an **interactive visualization library**

- Provides elegant, concise construction of versatile graphics

- Provides high-performance interactivity over very large or streaming datasets

- Can help anyone who would like to quickly and easily create interactive plots

How it works:

Not sure yet!

Demo:

- shows the usage of the Bokeh server

- UI and selection events can be processed to trigger more visual updates

https://demo.bokehplots.com/apps/crossfilter

Appeal:

- It does what we want!

- Great support / detailed user guide / plenty of examples and demos

Drawbacks:

- None of us have any experience with it.

- Looks a little bit complicated

Background:

- Need an **interactive interface for visualizing** which allows the user to select between different variables and explore their statistical relationship.

Applications…see next slide…

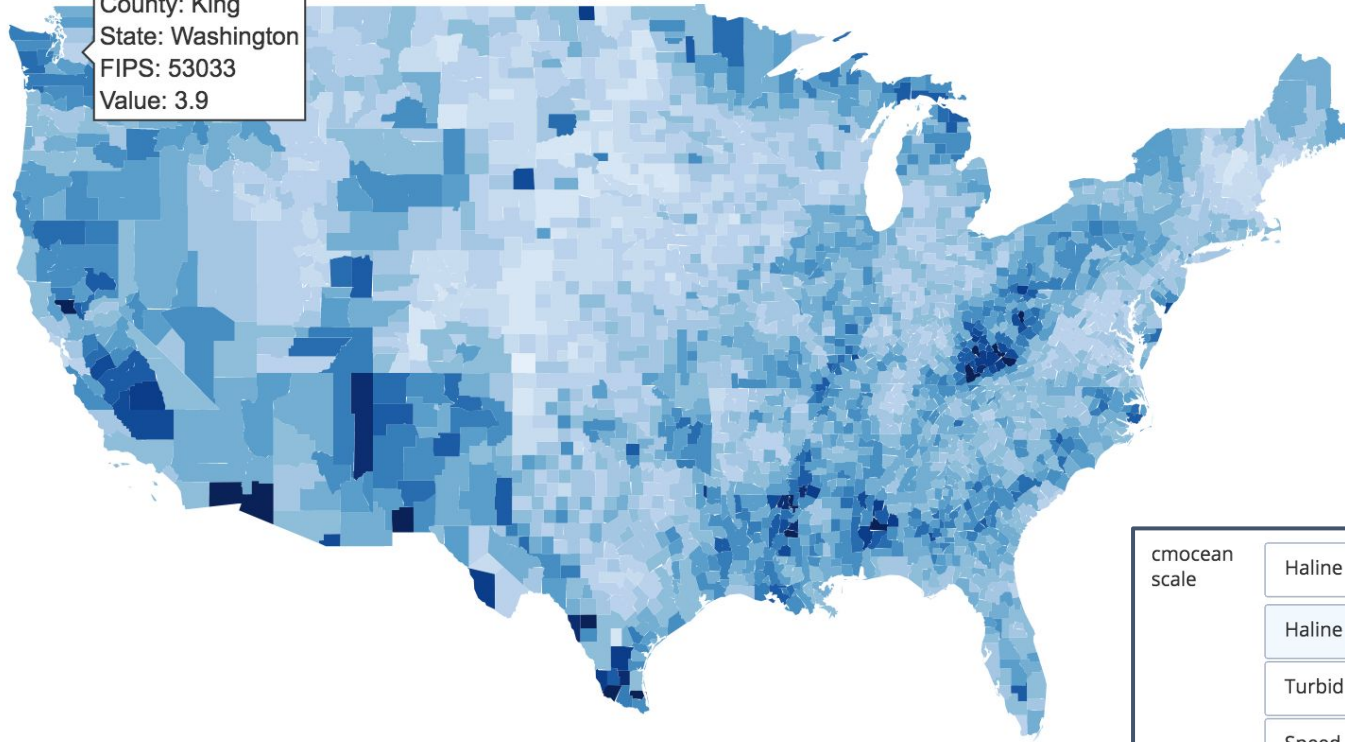Haven't really figured out the syntax / workflow for using yet.

USA by Unemployment %

County: King
State: Washington
FIPS: 53033
Value: 3.9

% unemployed

▬	> 12.0
▬	11.27 - 12.0
▬	10.53 - 11.27
▬	9.8 - 10.53
▬	9.07 - 9.8
▬	8.33 - 9.07
▬	7.6 - 8.33
▬	6.87 - 7.6
▬	6.13 - 6.87
▬	5.4 - 6.13
▬	4.67 - 5.4
▬	3.93 - 4.67
▬	3.2 - 3.93
▬	2.47 - 3.2

cmocean scale: Haline ▼   Trace type: 3d Surface ▼   Colorscale: Undo ▼   Lines: Show lines ▼

- Haline
- Turbid
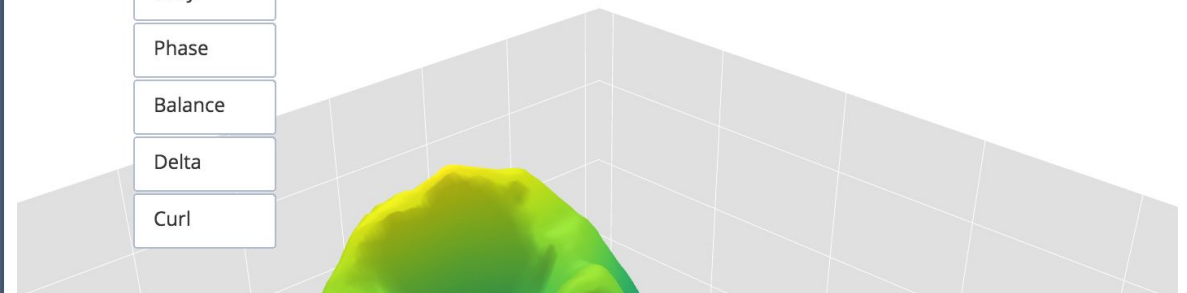- Speed
- Tempo
- Gray
- Phase
- Balance
- Delta
- Curl

180

160

Appeal:

- Has interactive features

- "If you **really want interactive plots** go with plotly over Bokeh. It offers a much greater level of interactivity than bokeh out of the box." ~ (some random reddit comment, 2017)

- Has examples of maps of census data for us to refer to

Drawbacks:
-Don't know how to pronounce

- Free version for public plots. **Private plots require a fee** and features are limited.

- **Maximum limit of 100 image exports** and chart saves per day in free version

- Apparently can be a little slow

- Relatively new library so there is not that much information available