Changing Cost of Living Across the United States

Jason Liu, Sanket Patel, Ye (Kevin) Pang

Project Concept

- While a dollar bill looks and feels the same all over the United States, its value is often very different depending on where it is spent
- Cost of living (CoL) is increasing throughout the United States but this rise may not be consistent everywhere
- Our project aims to analyze and visualize trends in the cost of various entities that influence the overall cost of living

Users

- General public - enables them to make informed decisions about budgeting for relocation or negotiating salaries to cover CoL expenditures adequately
  - Ex. Bob currently works in the suburbs of Georgia as a data scientist making $90,000 annually. His company has offered him a 15% raise to relocate to New York, NY. Can he expect an increase or decrease in net income after factoring in his rise and cost of living expenses in NYC?

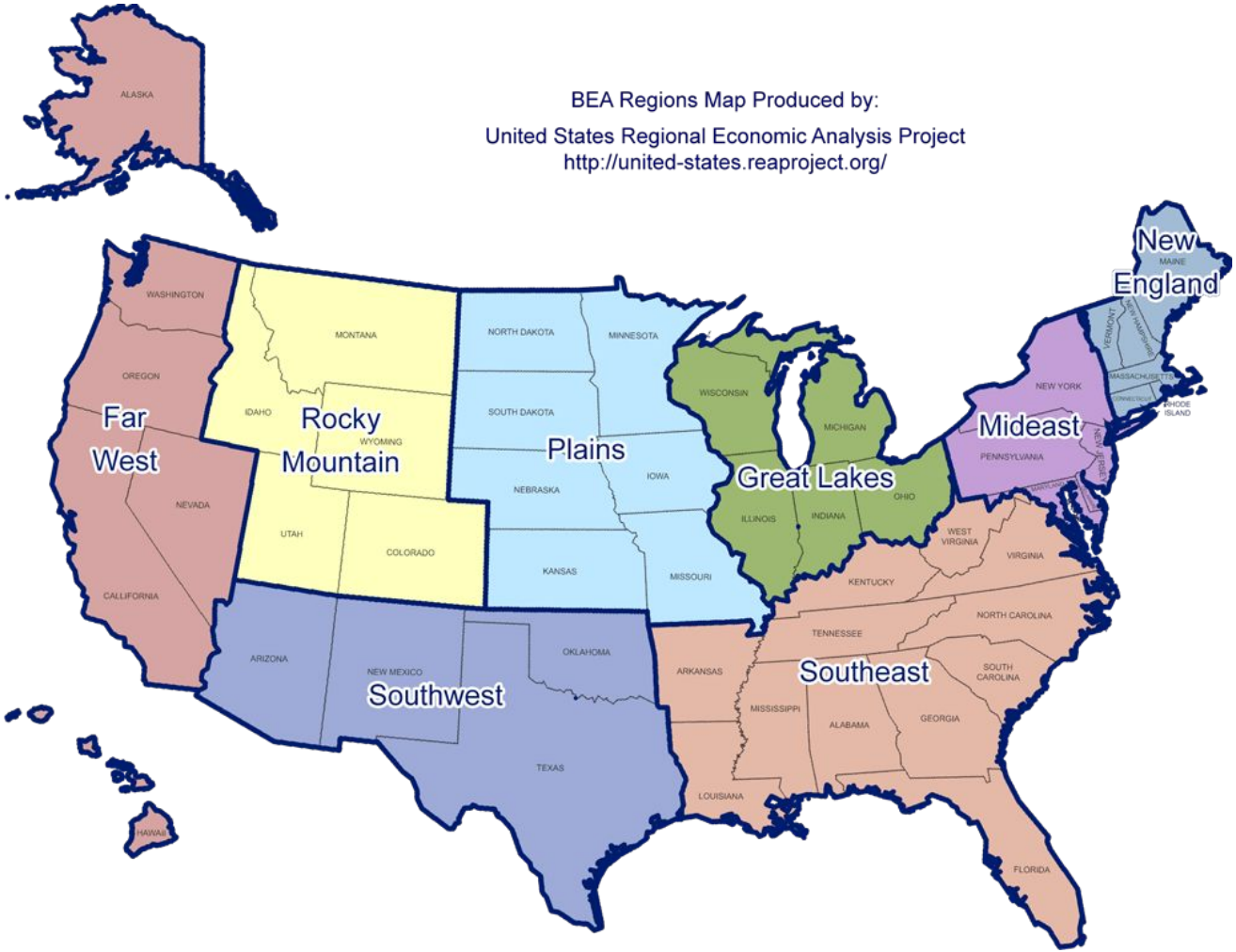
- Policy makers - enables them to better target cost-relief policies to those hardest hit
  - Ex. The government is considering a trial of a housing subsidy program to control rising prices. Where should the program be implemented to realize the greatest benefit?

Task #1: Explore CoL differences across the U.S. by:

- Per capita expenditures (selected categories)
- States and Regions
- Year

Data: Extracted regional per capita consumption expenditures in **tabular format** (download as csv file) from the Bureau of Economic Analysis (BEA) website.

- Preprocessed in **Python** to select relevant rows and perform any data munging necessary
- Expenditures are available in per capita and yearly time series from 1997 - 2017
- Per capita expenditures are in current dollars continuous integers
- Each row represents a time series for a state/region and a cost of living expenditure category
    - **Durable Goods** (1. Motor vehicles and parts, 2. Furnishings and durable household equipment, 3. Recreational goods and vehicles, 4. Other durable goods)
    - **Nondurable Goods** (1. Food and beverages purchased for off-premises consumption, 2. Clothing and footwear, 3. Gasoline and other energy goods, 4. Other nondurable goods)
    - **Services** (1. Housing and utilities, 2. Health care, 3. Transportation services, 4. Recreation services, 5. Food services and accommodations, 6. Financial services and insurance, 7. Other services, 8. Expenditures financed by third-party payers on behalf of households)
- Bureau of Economic Analysis also defines regions (multiple states) for comparison as well.


BEA Regions Map Produced by:
United States Regional Economic Analysis Project
http://united-states.reaproject.org/
ALASKA
HAWAII
WASHINGTON
OREGON
CALIFORNIA
NEVADA
IDAHO
MONTANA
WYOMING
UTAH
COLORADO
ARIZONA
NEW MEXICO
NORTH DAKOTA
SOUTH DAKOTA
NEBRASKA
KANSAS
OKLAHOMA
TEXAS
MINNESOTA
IOWA
MISSOURI
ARKANSAS
LOUISIANA
WISCONSIN
ILLINOIS
MICHIGAN
INDIANA
OHIO
KENTUCKY
TENNESSEE
MISSISSIPPI
ALABAMA
GEORGIA
FLORIDA
SOUTH CAROLINA
NORTH CAROLINA
VIRGINIA
WEST VIRGINIA
PENNSYLVANIA
NEW YORK
MARYLAND
DELAWARE
NEW JERSEY
MAINE
VERMONT
NEW HAMPSHIRE
MASSACHUSETTS
CONNECTICUT
RHODE ISLAND
Far West
Rocky Mountain
Plains
Great Lakes
Mideast
New England
Southwest
Southeast

Subtasks, Test Cases, Interactions For Task #1

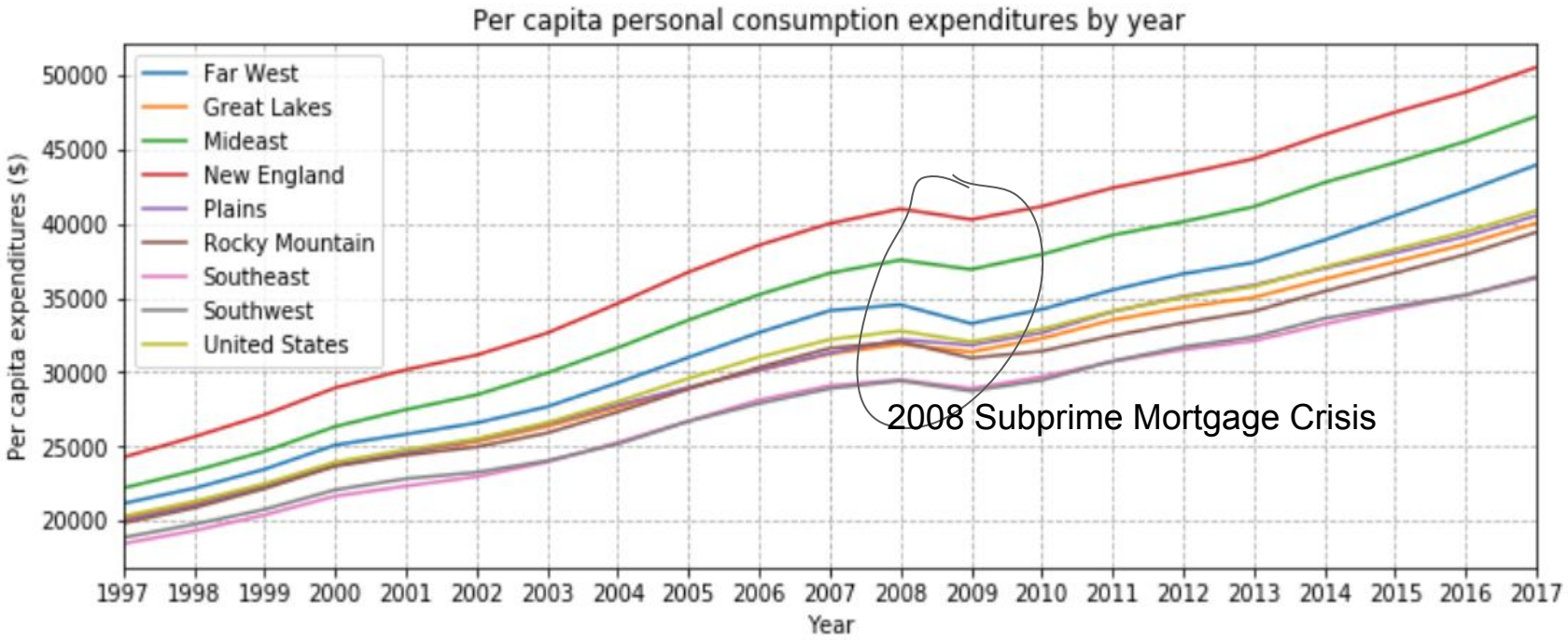
Subtasks & Test Cases

1. Compare state/region's rate of change and rank-order switches between categories
    - Which state has the sharpest change for `recreation services` between 1997 and 2017?
2. Reveal how states/regions show similar or different profiles across the categories
    - Are there any places that are similar to NY?
3. Show how large categories subdivides into subgroups, and which subgroups are bigger than others
    - Which subcategory under services is the largest in California and 2017?
4. Show over-time growth/contraction across the categories
    - Which year had the largest yearly increase for housing & utilities in New York?
5. Show proportion changes overtime across the categories
    - Can you find category(ies) that shrunk but re-expanded?

Interactions

- Peer comparisons (up to 3 - 5 selected regions/states)
- Linked charts
- Landing View + Highlighted/zoom in view for a selection
- Time lapse and rank movement animations
- Storyboards

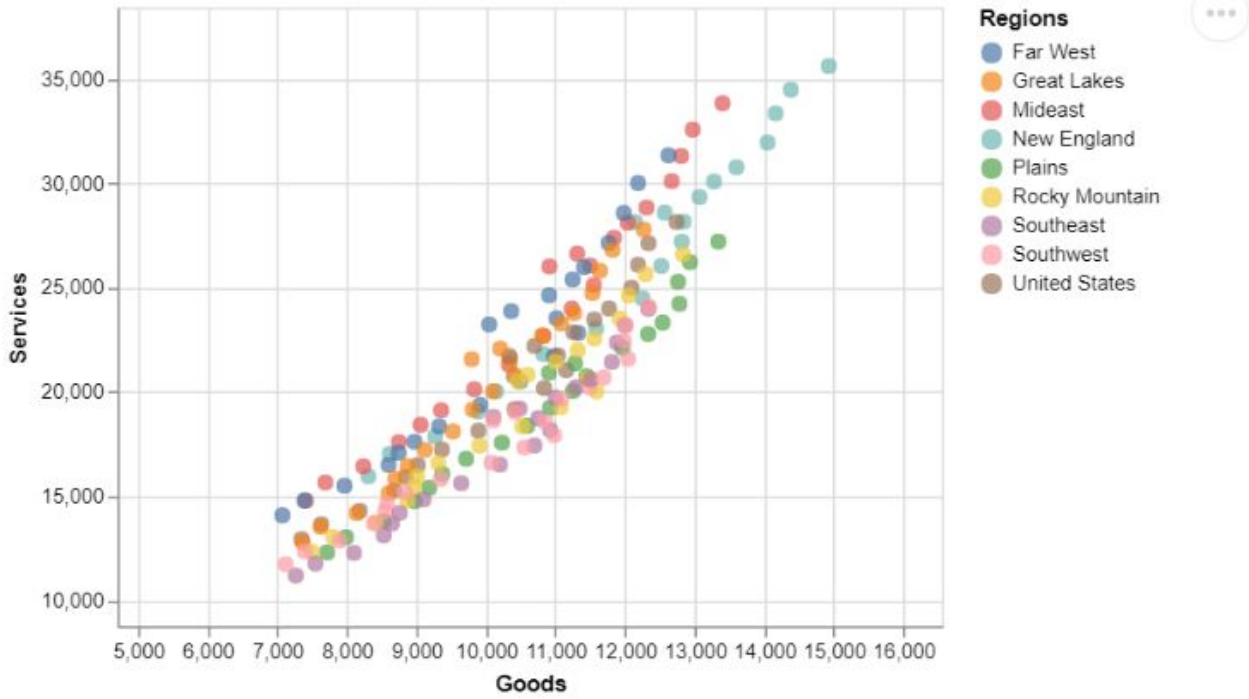
Viz Update For Task #1

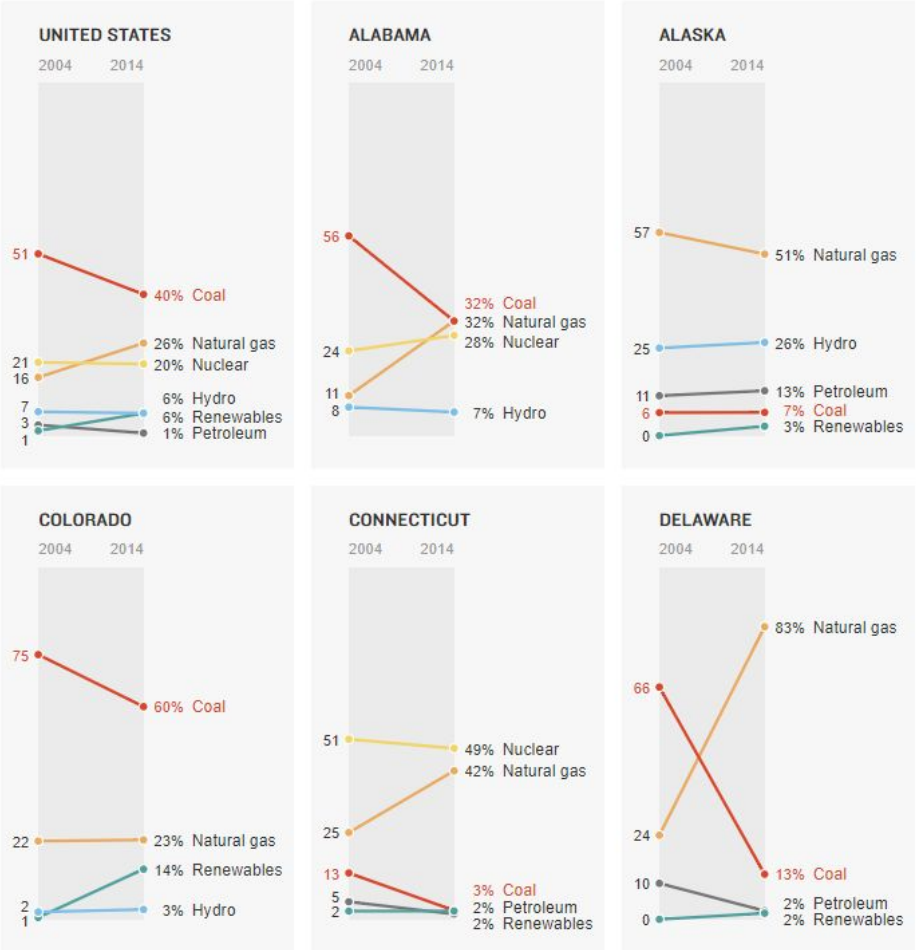


Plain boring line chart drawn in matplotlib…

- Use **Python** to data munge and change between long and wide formats
- Use **Altair** to speed up graphing code…
- Use **D3** to link charts and add interaction/animation

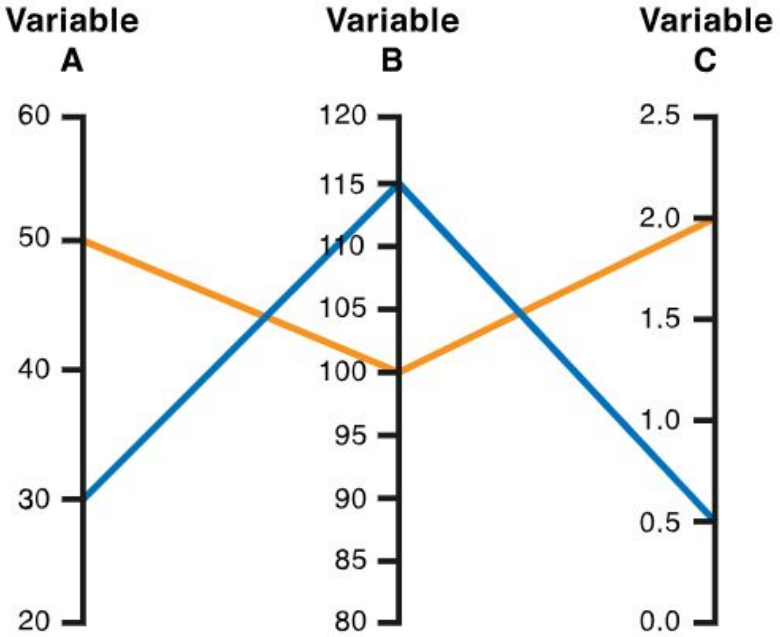
```
alt.Chart(data).mark_circle(size=60).encode(
    x='Goods',
    y='Services',
    color='Regions',
    tooltip=['Regions','Goods', 'Services', 'Year']
).interactive()
```





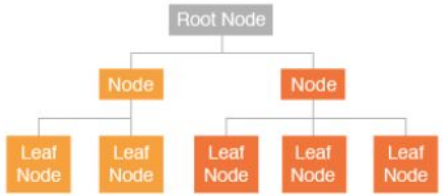
{Fake mockup, not using our data yet…}

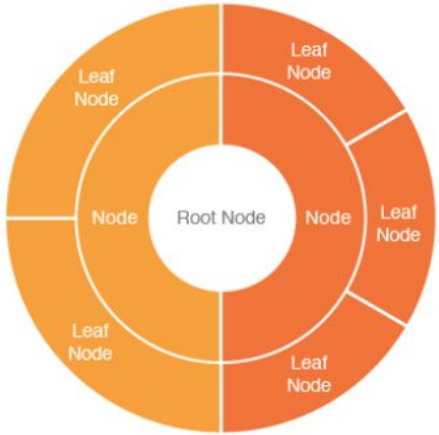
- Sub Task 1: Compare state/region's rate of change and rank-order switches between categories
- Insight: Certain regional categories exhibit sharp declines or increases
- Chart types considered:
  - Table with highlights
  - Bar graphs
  - Heatmaps
  - **Slopegraphs**
- Inspiration from
  - http://www.storytellingwithdata.com/blog/2018/6/1/swdchallenge-slopegraph
  - NPR's How Your State Generates Power

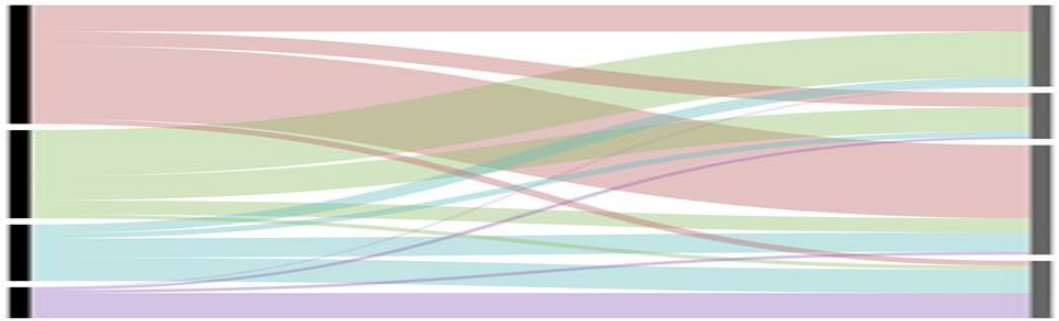


Data			
	Variable A	Variable B	Variable C
Item 1	50	100	2.0
Item 2	30	115	0.5

- Sub Task 2: reveal how states/regions show similar or different profiles across the categories
- Insight: a potential relocation candidate has these similar/dissimilar traits compared to current domicile state/region
- Chart types considered:
  - Table
  - Stacked bar chart
  - Treemap
  - **Parallel Coordinates**
- Inspiration from
  - https://datavizcatalogue.com/methods/parallel_coordinates.html



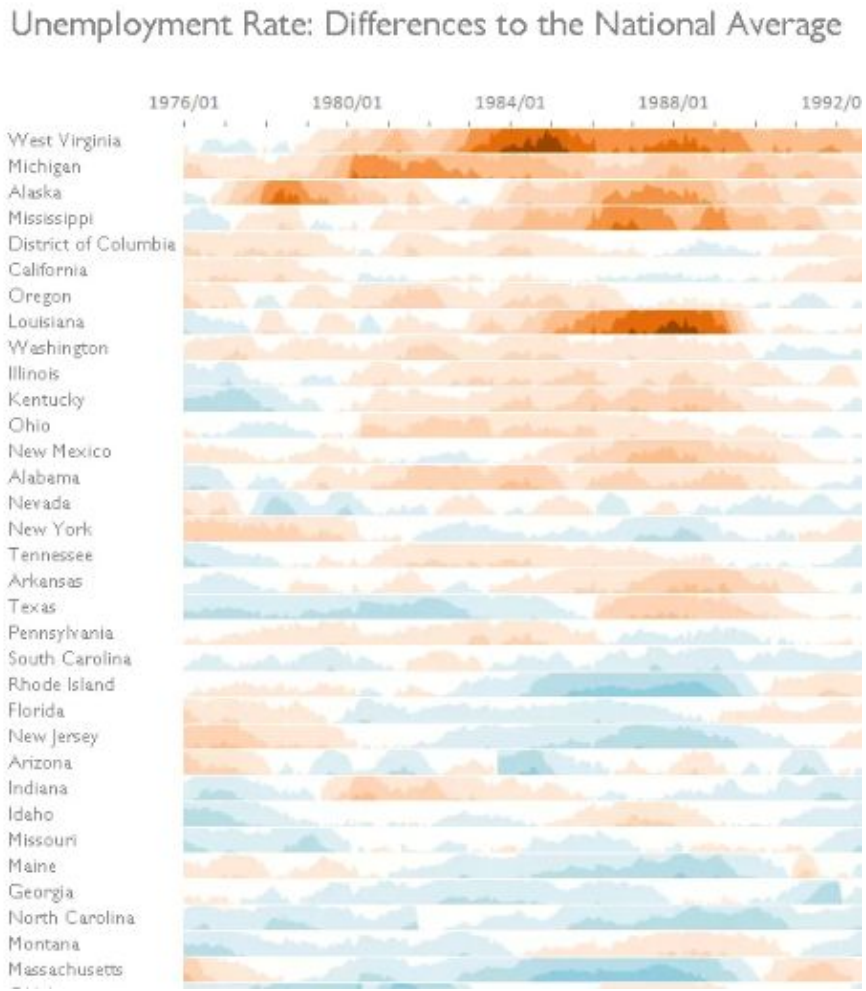




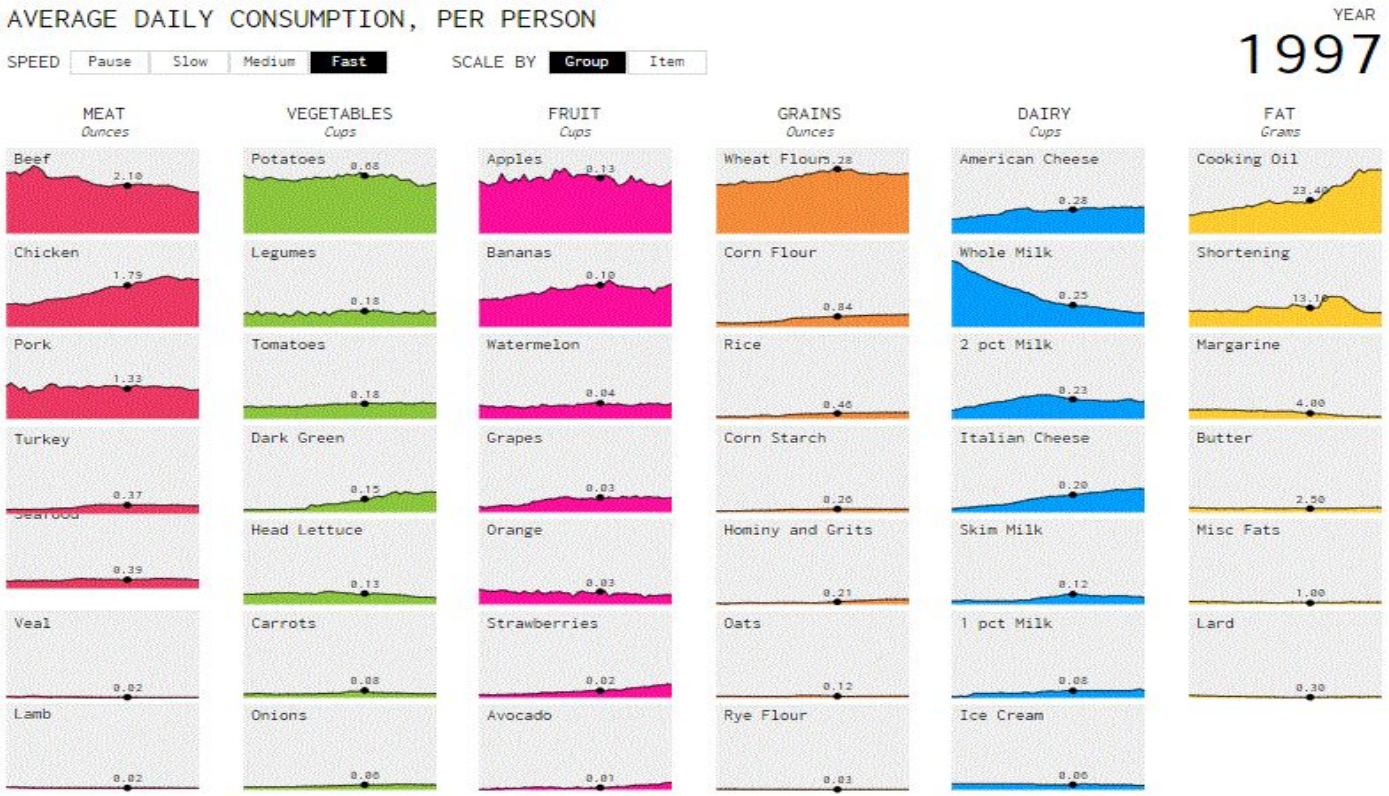
- <u>Sub Task 3</u>: show how large categories subdivides into subgroups, and which subgroups are bigger than others
- <u>Sub Task 5</u>: show proportion changes overtime across the categories
- Insights:
  - How much on average should I spend on ‘recreational services’ in the Rocky Mountain?
  - What make the biggest expenditures before and now?

Chart types considered:

- **Sunbursts**
- Stacked bar chart
- Treemap
- Sankey
- **Alluvial Diagrams**

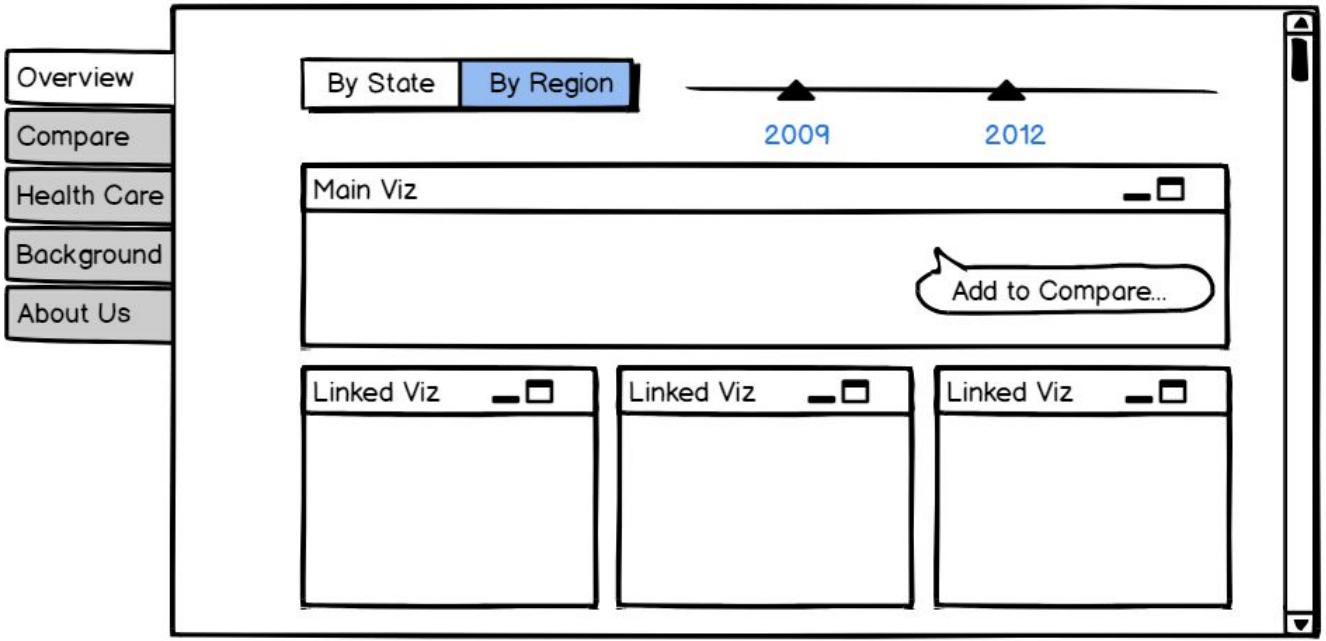


- Sub Task 4: show over-time growth/contraction across the categories
- Insight(s):
  - Which expenditure items are growing or shrinking the fastest in an state/region?
  - How does the rate change compare to the rest of U.S?
- Chart types considered:
  - Table
  - Line chart
  - Diverging bar chart
  - **Horizon chart**
- Inspiration:
  - https://excelcharts.com/horizon-graph-reorderable-matrix-unemployment-rate-1976-2012/

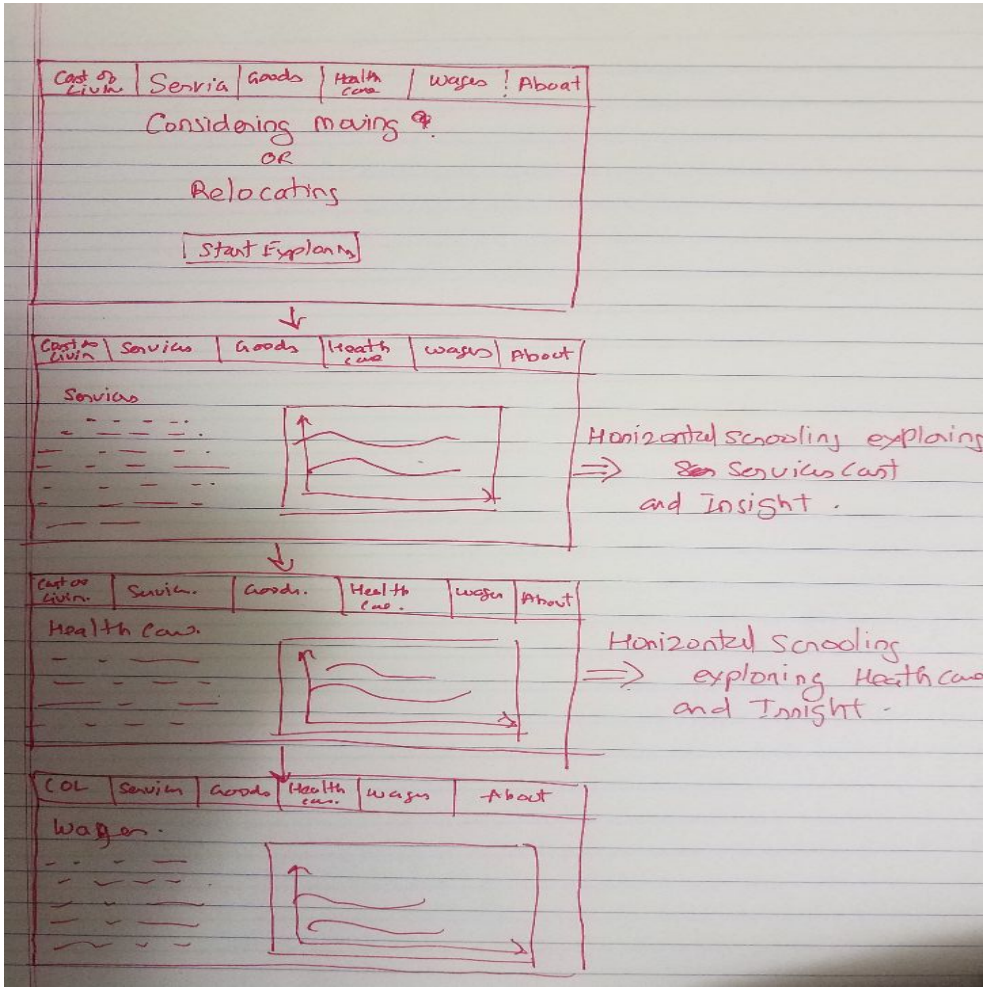


Time Lapse Animation (will model after this example)

Changing Cost Of Living Across The United States



- Web app? (Undecided)



- Storyboards? (Undecided)

Task 2:

Task: Explore and compare trends in cost of living and wages across the united states for the last 10 years.

Data: Annual wage data exported from Bureau of Labor Statistics

Subtasks:

- Where are wages really growing and where are wages not growing?
- Which occupation has seen the most growth over the last 10 years?
- Which states have seen the most increase in Consumer Price Index over time?
- Which states are the top candidates for Tech and Teaching jobs by wages?
- Best state for a given occupation to maximize savings.

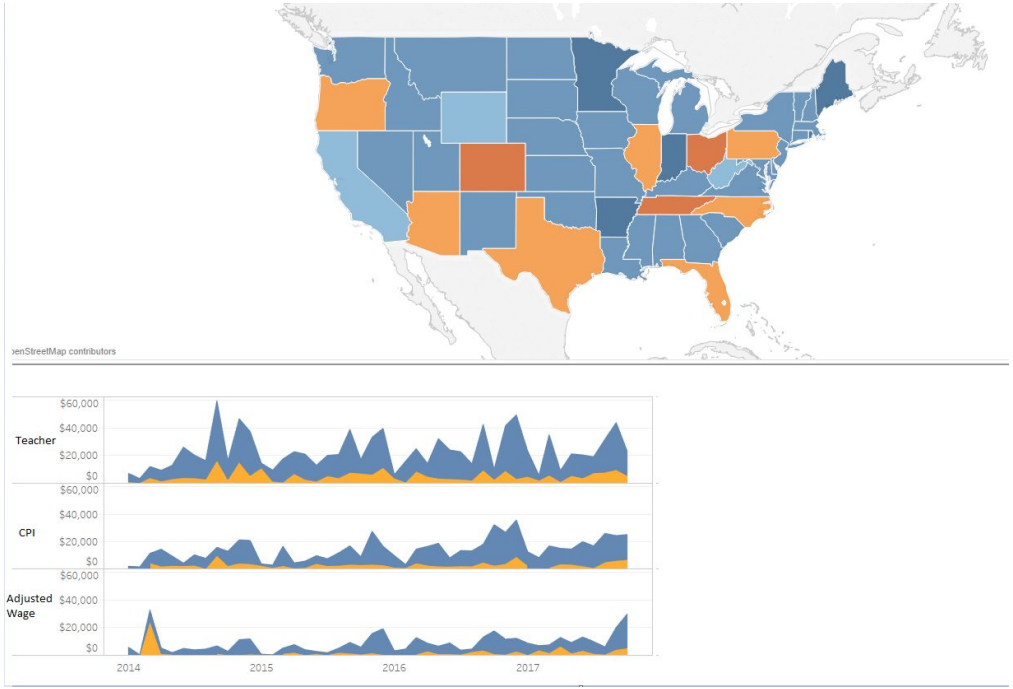
Insight: User looking for a job in Software development finds that washington is a better option than california even though average salary might be higher there.

Testing:

- Performance Testing
- Task based testing from general public (fellow MIDS students and friends)

Iteration 1

- Decided to use D3 and altair for visualization any python for data processing
- Variable to visualize
  - ~900 occupation types
  - 50 states
  - Wages per occupation
  - Per state
  - Consumer Price Index per state
  - Changes in wage and CPI over time
  - Correlation
- Initial Viz included a map of US linked to couple of line charts and dropdown to select occupation



Issues:

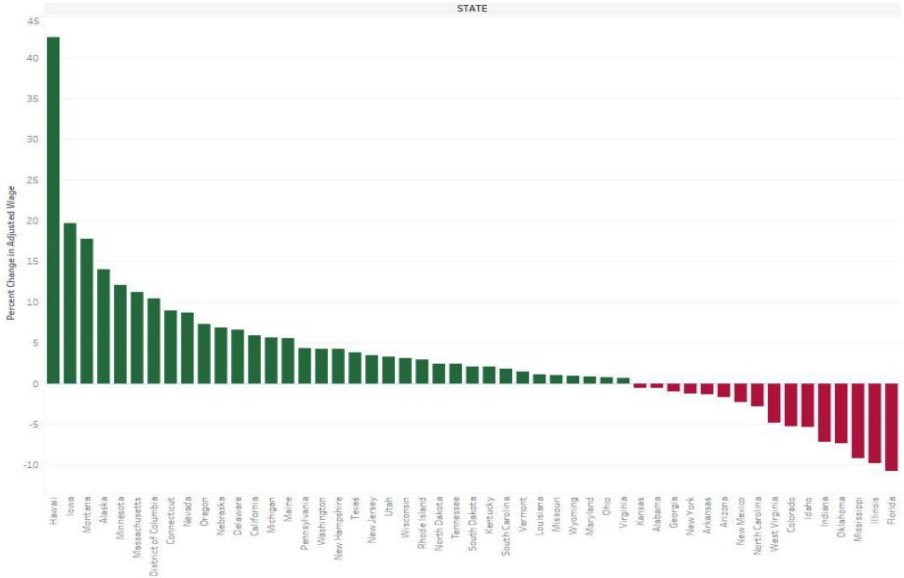
- Map give a good overview at a glance but difficult to figure out ranking

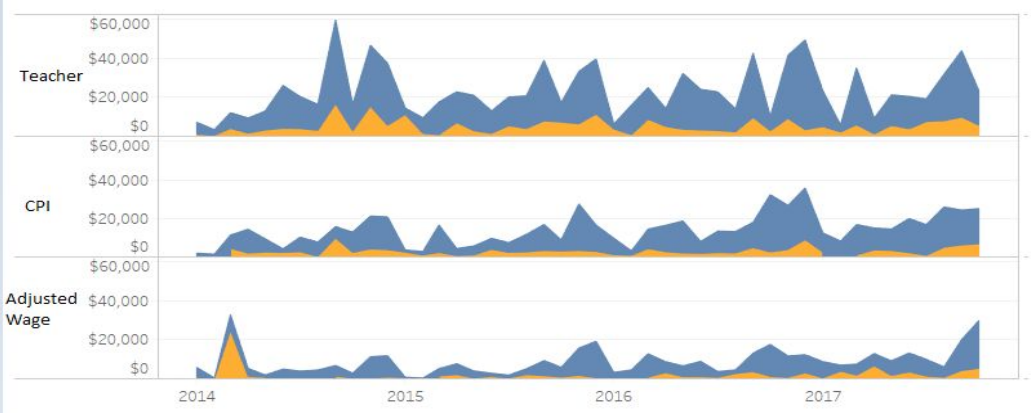
Iteration 2

- Ability to see ranking of states with most growth for given occupation
- Allow use to select a occupation through a drop down and range of date using a slider
-

Issues/Missing:

- Still missing ability to explore correlation between wage and CPI



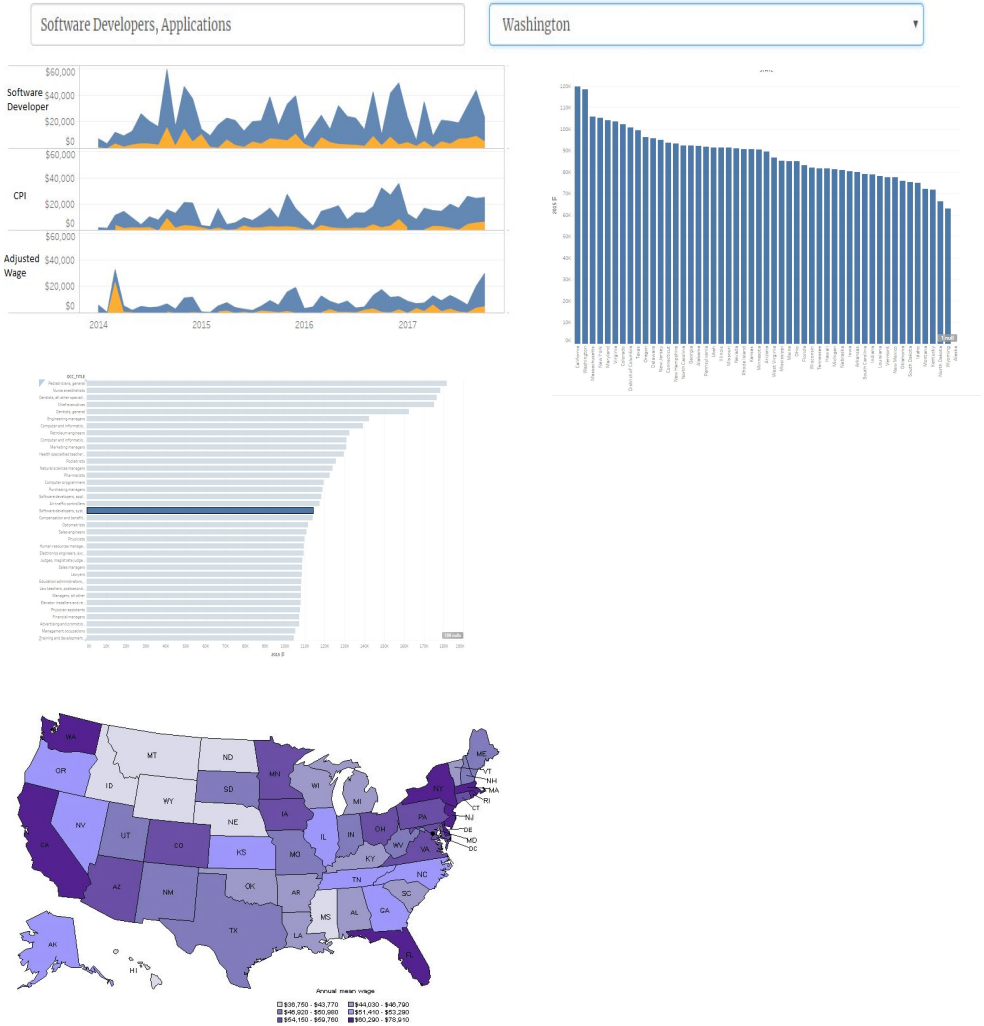


Iteration 3

- An overview shows the map of US with average wage growth over the last 10 years encoded as colors
- Another graph shows the ranking of Occupations based on increase in wage.
- Selecting a wage highlights the state that has seen the most increase and other way
- User selects Occupation and State from a dropdown
- This triggers a line chart that shows the trends in wage and CPI over time for that state.
- Another graph shows the comparison to other states adjusted for cost of living
- A chart to show wages for other occupation in the same state

Testing:

- Performance Testing
- Task based testing from general public (fellow MIDS students and friends)



Task 3: Out of Pocket Healthcare Spending

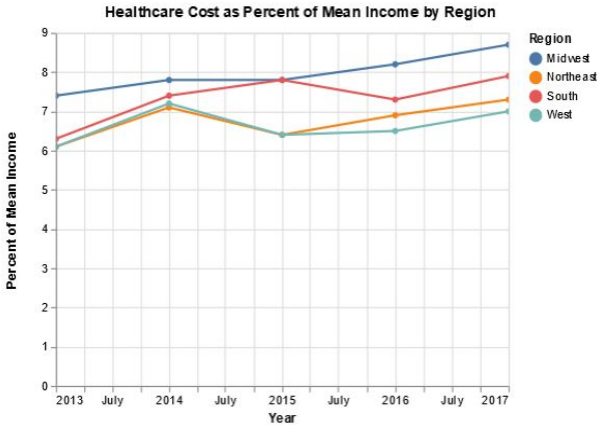
- Data: Bureau of Labor Statistics Consumer Expenditure Survey
- Task: Summarize how out of pocket healthcare costs have increased over a 5 year period (2013-2017)
  - By region
  - By age group
  - By overall population
  - By cost type
- Subtasks
  - Identify the regions and age group(s) that spend the most on out-of-pocket healthcare expenses
  - Compare different types of healthcare spending among age groups
  - Compare trends of healthcare spending by region and age group

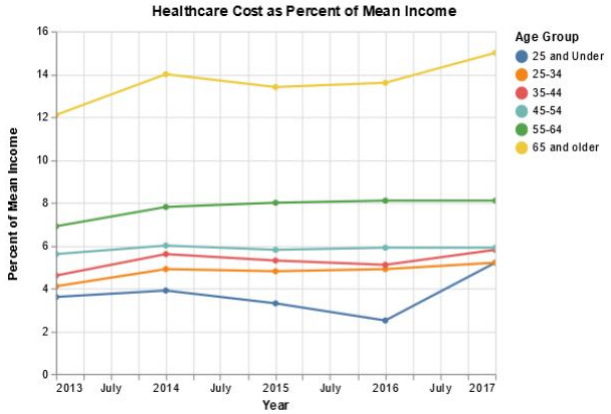
Task 3: Out of Pocket Healthcare Spending

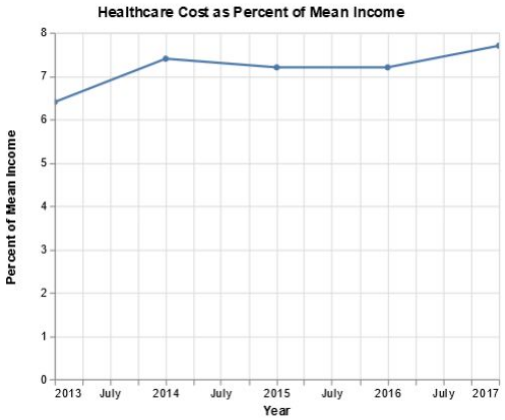
- Users
  - Policy makers
  - General public
- Insight: Knowing what kind of healthcare costs place the greatest burden on the population better informs the design of public programs to control those costs.
  - Ex. A congresswoman is developing legistation to help vulnerable populations manage their healthcare costs and would like to know what kind of people are spending the most on healthcare and what those types of costs are.
- Planned technologies to use
  - Altair in python
  - Use d3 as necessary for interaction implementation

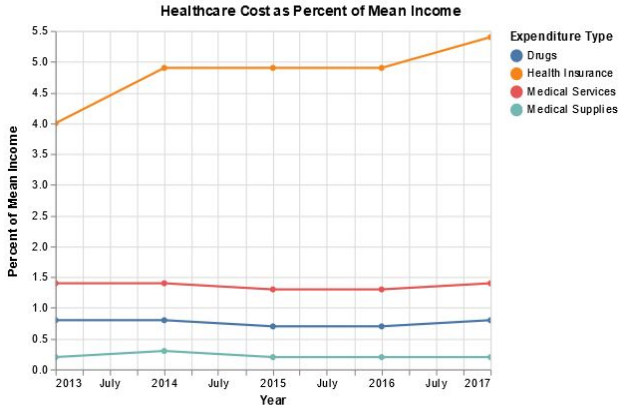
Task 3: Visualization Mock-ups

- Multiline charts
- Good for comparing overall trends by different categories but it is difficult to make comparisons on a secondary variable



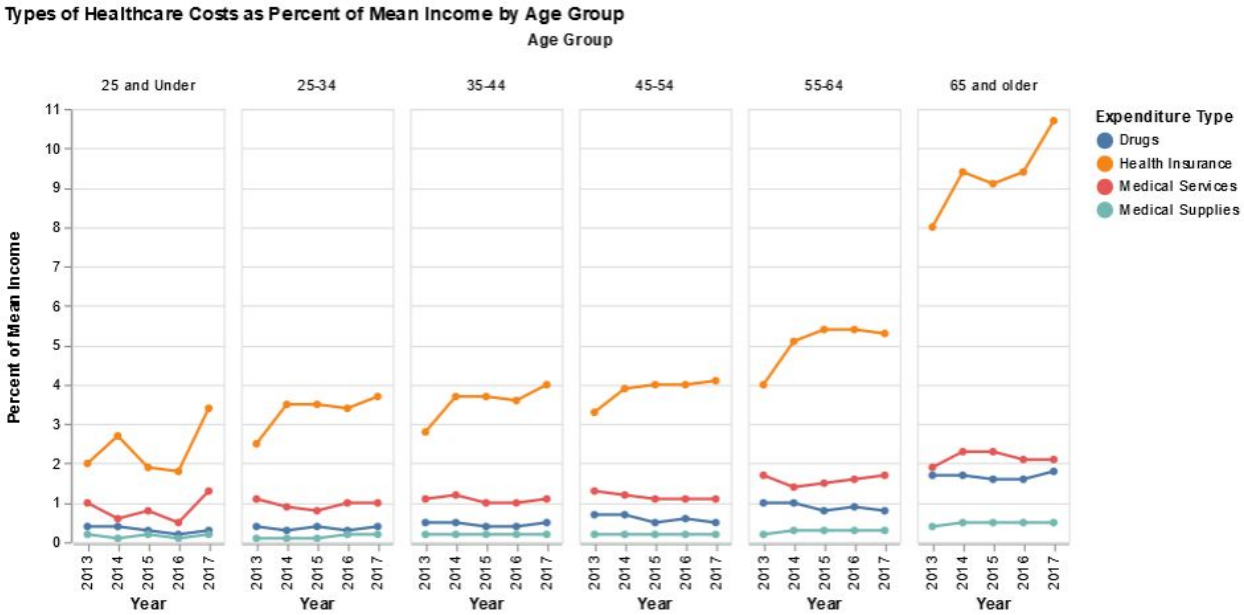






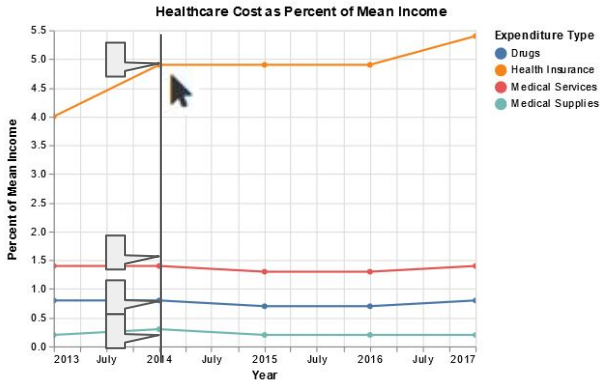
Task 3: Visualization Mock-ups

- Comparison of various types of healthcare costs by age group
  - Faceted design to allow for easy comparison among age groups

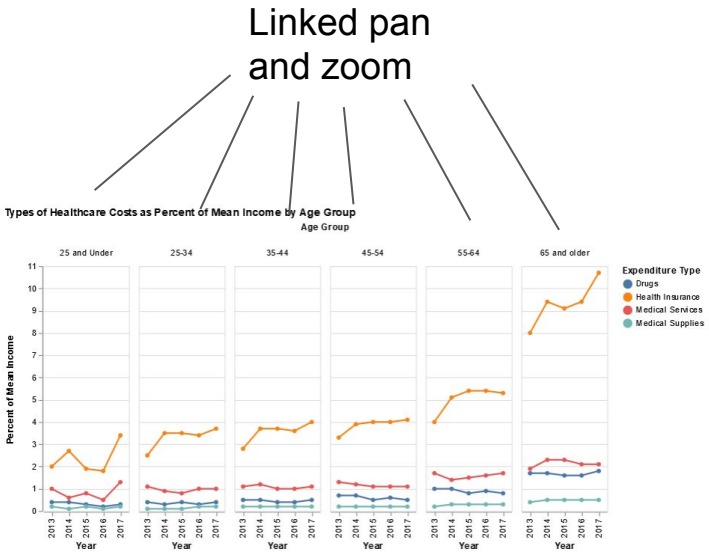


Task 3: Planned Interactions

- Multi-line tooltip on mouse-over for easy comparison among categories



- Linked pan and zoom across different facets



Task 3: Sample Testing Plan

Task List:

- In 2017, what was the greatest type of healthcare expenditure made by a person 65 years or older?
- Which region spent the greatest percentage of their income on healthcare expenditures in 2017?
- What percentage of their income would a typical 37 year old spend on health insurance in 2015?
- What age group spends the most overall on healthcare?
- What percentage of income can a 60 year old expect to spend on drugs?

Questions

Please send us your feedback to:

yp2182@berkeley.edu