

# Data Visualization: Consumer Price Index (CPI)

<b>Overview.....</b>	<b>3</b>
Data Source.....	3
Key Attribute and Dimensions in the Data:.....	3
Goals.....	3
Tasks.....	4
<b>Solution.....</b>	<b>5</b>
Screenshots.....	5
Control.....	5
Test.....	5
Links to Visualization.....	6
<b>Summary of Design.....</b>	<b>6</b>
Design Element 1: Sorted Based Percentage Change.....	6
Design Element 2: Limited to Top 10.....	6
Design Element 3: Reference Lines.....	6
<b>Visualization Evaluation.....</b>	<b>6</b>
Overview.....	6
Approach.....	7
Procedure (Methodology).....	8
People Recruited.....	8
Results.....	9
Table of Results Collected During the Experiment.....	9
Calculated Results for Each Chart Type.....	9
Conclusion.....	9
<b>Findings.....</b>	<b>10</b>
What Worked Well.....	10
What Should be Refined.....	10
<b>Appendix A: Part 1: Finding your Data.....</b>	<b>11</b>
The Data Source.....	11
Key Attribute and Dimensions in the Data:.....	11
Goals for Working with that Data.....	11
Identify Existing Relevant Visualizations.....	12
Example 1: Table of Consumer Price Index Statistics.....	12
Example 2: Select Category Bar Chart for 1 Month Percentage Change.....	13
Example 3: Select Category Bar Chart for 1 Month Percentage Change.....	13
Other Charts Not Included in Writeup.....	14
Example A: CPI-U News Release Companion File.....	14
Example B: Relative importance, and over-the-year and over-the-month percent change.....	15
<b>Appendix B: Part 2: Sketching your Data.....</b>	<b>16</b>
Recap of Goals Initially Identified.....	16
Task 1: Citizens Analyzing CPI for Understanding and Relevance.....	16
Task 2: Policy Makes Analyzing CPI Influencers.....	17
Task 3: CPI Changes MoM.....	18

**Appendix C: Part 3: A Plan for Evaluation.....18**

    Target Question..... 18

    Recruited People..... 19

    Measures and Outcomes..... 19

    Approach..... 19

    Methodology.....20

    Evaluation Criteria.....20

**Appendix D: Part 4: Final Project..... 21**

    Develop Solution..... 21

    Recap.....21

# Overview

## Data Source

On a monthly basis, the Bureau of Labor Statistics within the Department of Labor posts a monthly news release on changes to the Consumer Price Index (CPI). An example of the report can be found here: <https://www.bls.gov/news.release/pdf/cpi.pdf>). Along with the news release, they update the data source and a selection of charts to accompany the news release.

The data source can be found here: **CPI-U News Release Companion File:** <https://www.bls.gov/web/cpi/cpi-u.xlsx>

**Raw Versions** of the data and **Attribute Reference Data** can be found here: <https://download.bls.gov/pub/time.series/cu/>

## Key Attribute and Dimensions in the Data:

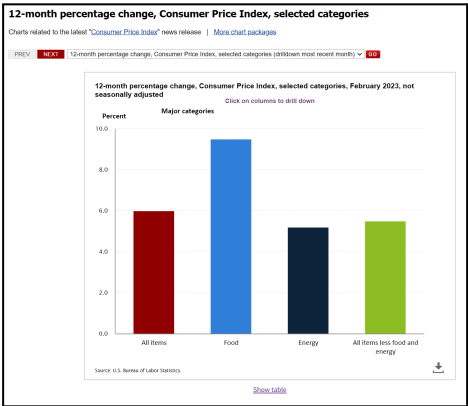
- **Item Code** - Identifies items for which data observations pertain.
- **Item Name** - Full names of items.
- **Item Code Parent** - How a specific item rolls-up into a larger grouping or category as the CPI is calculated
- **Item Code Level** - The level of the item code hierarchy where the item code is located
- **Relative Importance** - the weight associated with the item code (all item codes at a specific level will sum to 100)
- **Unadjusted Percent Change (Twelve Months)** - The percentage change for the specific item code when comparing the current month to the same month 1 year (12 months) ago.
- **Unadjusted Effect on All Items (12 Months)** - The contribution of the specific item to the CPI for All Items when comparing the current month to the same month 1 year (12 months) ago. This can be a positive or negative contribution to the overall CPI. This field, in conjunction with the Relative Importance can be used to compute a CPI for a subset of items.

## Goals

While the CPI news release and charts are thorough, they focus on the top level CPI number. This is effective at describing the why behind inflation, but given the amount of aggregation, this package falls short of successfully communicating the “so what” of changes in CPI and inflation to the average person.

Where the Bureau of Labor Statistics provides charts, they tend to fall short. Either they are way too detailed (tables) or way too general (bar chart).

Table 1. Consumer Price Index for All Urban Consumers (CPI-U): U.S. city average, by expenditure category, February 2023 1982=100-100 (unless otherwise noted)									
Expenditure category	Index Feb. 2023	Unadjusted percent change		Unadjusted percent change		Seasonally adjusted percent change		Seasonally adjusted percent change	
		Jan. 2023	Jan. 2022	Feb. 2023	Feb. 2022	Jan. 2023	Jan. 2022	Feb. 2023	Feb. 2022
All items	100.000	283.716	259.170	300.840	260.000	0.0	0.0	0.0	0.0
Food	13.021	292.764	270.736	300.000	260.000	0.0	0.0	0.0	0.0
Food at home	8.727	274.668	261.455	300.000	260.000	0.0	0.0	0.0	0.0
Cereals and bakery products	1.106	306.193	304.254	300.000	260.000	0.0	0.0	0.0	0.0
Meats, poultry, fish, and eggs	1.840	301.240	302.737	301.803	260.000	0.0	0.0	0.0	0.0
Dairy and related products	1.613	242.986	272.042	302.271	260.000	0.0	0.0	0.0	0.0
Fruits and vegetables	1.008	332.675	301.039	301.249	260.000	0.0	0.0	0.0	0.0
Nonalcoholic beverages and tobacco	1.046	192.210	213.338	274.962	260.000	0.0	0.0	0.0	0.0
Other food at home	2.344	297.149	264.746	304.602	260.000	0.0	0.0	0.0	0.0
Food away from home	2.292	300.000	295.077	300.000	260.000	0.0	0.0	0.0	0.0
Energy	7.076	267.771	283.230	281.673	260.000	0.0	0.0	0.0	0.0
Energy commodities	1.004	211.020	205.243	206.280	260.000	0.0	0.0	0.0	0.0
Fuel oil	1.147	344.179	405.545	418.039	260.000	0.0	0.0	0.0	0.0
Natural gas	1.344	267.402	267.116	266.169	260.000	0.0	0.0	0.0	0.0
Gasoline (all types)	3.246	300.000	284.759	286.774	260.000	0.0	0.0	0.0	0.0
Energy services	1.502	230.017	272.042	280.263	260.000	0.0	0.0	0.0	0.0
Electricity	2.379	236.403	264.628	266.487	260.000	0.0	0.0	0.0	0.0
Utility-gas and water	1.641	236.403	264.628	266.487	260.000	0.0	0.0	0.0	0.0
All items less food and energy	70.493	284.009	301.942	304.011	260.000	0.0	0.0	0.0	0.0
Commodities less food and energy	21.203	144.009	165.540	166.246	260.000	0.0	0.0	0.0	0.0
Alcohol	1.000	127.000	127.000	130.000	260.000	0.0	0.0	0.0	0.0
New vehicles	4.200	180.000	177.000	177.000	260.000	0.0	0.0	0.0	0.0
Used cars and trucks	1.000	210.000	180.000	180.000	260.000	0.0	0.0	0.0	0.0
Medical care	1.000	284.000	284.000	284.000	260.000	0.0	0.0	0.0	0.0
Alcoholic beverages	1.043	268.128	262.284	262.377	260.000	0.0	0.0	0.0	0.0
Medical care excluding products	1.000	124.000	124.000	124.000	260.000	0.0	0.0	0.0	0.0
Services less energy	58.074	302.072	301.000	301.071	260.000	0.0	0.0	0.0	0.0
Shelter	36.903	344.000	344.000	344.000	260.000	0.0	0.0	0.0	0.0
Rent of primary residence	7.521	354.000	354.000	354.000	260.000	0.0	0.0	0.0	0.0
Owners' equivalent rent of residence	29.382	339.000	339.000	339.000	260.000	0.0	0.0	0.0	0.0
Medical care services	1.000	284.000	284.000	284.000	260.000	0.0	0.0	0.0	0.0
Physician services	1.000	284.000	284.000	284.000	260.000	0.0	0.0	0.0	0.0
Medical equipment	1.000	284.000	284.000	284.000	260.000	0.0	0.0	0.0	0.0
Transportation services	1.000	284.000	284.000	284.000	260.000	0.0	0.0	0.0	0.0
Motor vehicle maintenance and repairs	1.100	331.136	321.780	322.023	260.000	0.0	0.0	0.0	0.0
Motor vehicle insurance	0.208	444.000	444.000	444.000	260.000	0.0	0.0	0.0	0.0
Arts and recreation	0.274	222.000	222.000	222.000	260.000	0.0	0.0	0.0	0.0



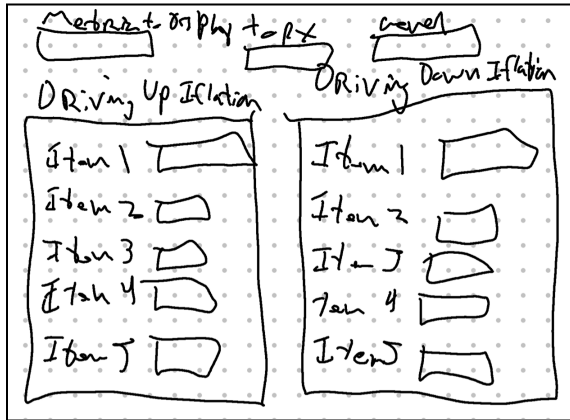
Something with more focus and less noise in the visualization is needed.

To address this need or citizens and researchers, a chart needs to be created in the middle of these two bookends, focused on the addressing the following questions and opportunities:

- What is driving the month over month changes?
- Are there offsetting increases and decreases that are obscured in the highly aggregated metric?
- How can I enable a citizen to accurately and efficiently identify expenditure categories driving the CPI number?

## Tasks

Create a new visualization that drive efficiency (reduced clicks and time saved) as well as improved accuracy for **Policy Makers and Citizens Analyzing CPI Influencers.**



### Why is a task pursued? (goal)

Understand what is primarily driving CPI changes (inflation) in the current period.

### How is a task conducted? (means)

Review what the top x influencers are to CPI in the positive and negative direction.

### What does a task seek to learn about the data? (characteristics)

Use outcomes to understand what is driving the CPI changes (winning and losing) and where to focus efforts to combat inflation.

### Where does the task operate? (target data)

Operate with the target data of Consumer Price Index dataset from the Bureau of Labor Statistics.

### When is the task performed? (workflow)

Perform monthly to understand changes in pricing and make better policy choices to focus effort.

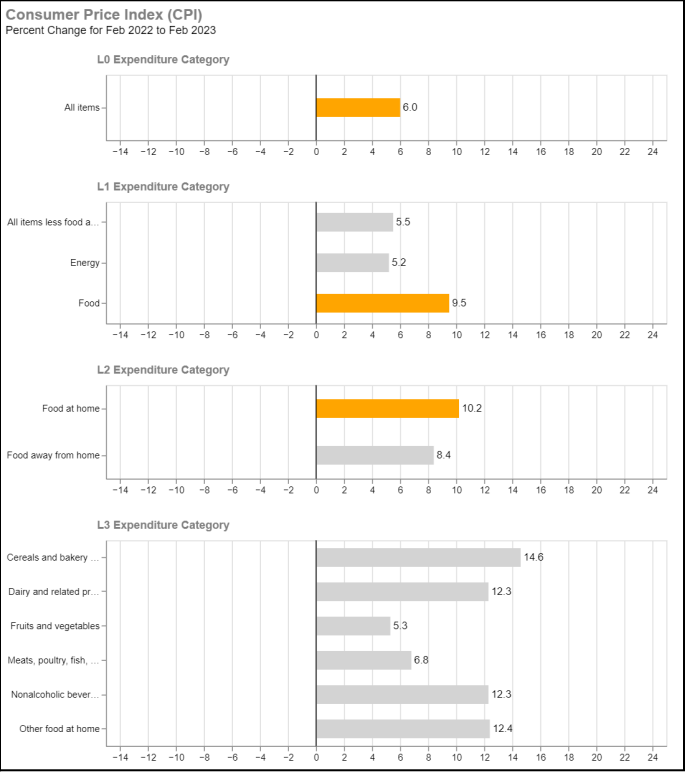
### Who is executing the task? (roles)

Economists or policy making that can influence prices or consumer behaviors.

# Solution

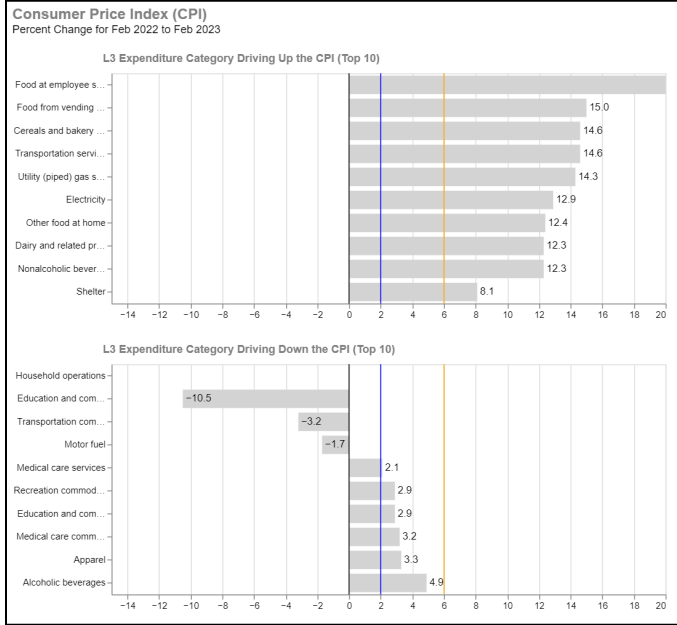
## Screenshots

### Control



The chart used for the control chart in the experiment is a drill down chart similar to what was created to the Bureau of Labor Statistics as a companion for the CPI press release ([here](#)). It shows 4 levels of the Expenditure Categories. At the top is the Level 0 (L0) which is the All Items category. Changes to this category are approximately equal to changes to the CPI and changes to CIP are inflation. As the user selects bars, they filter the child category (level 1 through level 3). These show more granular representations. Each bar represents the percent change in the CPI for that category and is approximately equal to inflation for that category.

### Test



The chart used for the test chart in the experiment is a top x chart, where the user sees the top 10 largest positive changes to the Expenditure Category and the top 10 smallest (sometimes negative) changes to the Expenditure category. This chart has actually removed the interactivity and is focused on ease of use and speed. It enables a user to easily distill the information related to influence on the overall CPI and ignore values close to the All Items CPI that aren't moving the score in either direction. A blue line is added to indicate the target CPI and an orange line is added to indicate the All Items average on the chart that is more granular.

## Links to Visualization

An interactive version of both charts are hosted on google drive and accessible through the links below. If you would like to explore the files, please download them to your computer. Once on your computer, open the file in chrome, and the visualization will render in a browser tab from the html file.

**Link to Control Visualization File:** [CONTROL - L3\\_final\\_combo.html](#)

**Link to Test Visualization File:** [TEST - L3\\_top\\_final\\_combo.html](#)

## Summary of Design

The following section summarizes key elements of the design and provides accompanying justification for design decision.

### Design Element 1: Sorted Based Percentage Change

The test chart is in sorted order based on the percentage change. The chart showing expenditure categories driving up the CPI is sorted in descending order, and the chart showing expenditure categories driving down the CPI is sorted in ascending order. This enables the user to see the most influential expenditure category at the top of both charts. It also prevents the user from needing to search through multiple charts and parent categories to find the top items. This design element should save the user time and reduce searching through an interactive chart.

### Design Element 2: Limited to Top 10

The test chart utilizes a top k expenditure categories feature. In the case of the chart used in the test, we see the top 10 largest and top ten smallest expenditure category records (in 2 separate charts). This allows the user to focus on the most extreme values that are pulling the CPI in opposite directions. By limiting the records, we are able to reduce a lot of the noise in the chart, place all the records on one screen, and increase the speed of the user to find the top items.

### Design Element 3: Reference Lines

The test chart utilizes reference lines to indicate the All Items average at 6.0% (known as the CPI). It also utilizes a reference line to indicate the Target CPI in the United States. By including these reference lines, the user can draw inferences from the chart. For example, it helps the user see that there are at least 10 expenditure categories driving down the CPI, however most of them are still above the Target CPI.

## Visualization Evaluation

For the evaluation, I will use a Summative Experimental Evaluation as a Quantitative Evaluation. I will use a within participants factor where each person sees all conditions. Given the medium (likely a local copy) I will be present to manually measure the dependent variables and collect data. I will also be able to collect any qualitative data from the users from Thinkalouds that follow the experimental task.

# Overview

The following sections provide an overview of the visualization evaluation design and structure.

## Target Question:

- **Question:** What CPI representation lets people most accurately determine the top X and bottom X influencers to CPI at a given level of items.
- **Hypotheses:** A sorted top x and bottom x plot will be most accurate.

## Measures and Outcomes:

We would look at 3 key measures to determine the quality of the solution to answer our core questions:

- **Time:** How long does it take the user to complete the given task? (lower is better)
- **Accuracy:** How Accurate can people complete a given task? (higher is better)
- **Number of Interactions:** How many steps does it take to complete the given task? (lower is better)

# Approach

For the evaluation, I will use a Summative Experimental Evaluation as a Quantitative Evaluation. I will use a within participants factor where each person sees all conditions. Given the medium (likely a local copy) I will be present to manually measure the dependent variables and collect data.

## Experimental Tasks:

[Comparison] Determine the top 3 influential segments for CPI (either direction):

- Determine the top 3 segments increasing CPI for the month
- Determine the top 3 segments decreasing CPI for the month

## Experimental Stimuli:

- **Control:** Recreation of the BLS Bar Chart of 12-month percentage change, Consumer Price Index
- **Test:** New bar chart layout focused on top X influencers to increase CPI and top X influencers to decrease CPI

## Independent Variables:

- Configuration of Bar Chart to explore and display dataset (layout, filtering, and features).

## Dependent Variables:

- **Time:** How long does it take the user to complete the given task?
- **Accuracy:** How Accurate can people complete a given task?
- **Number of Interactions:** How many steps does it take to complete the given task?

## Control Variables:

- **Experimental Stimuli:** Pseudo-randomize order of visualizations (experimental stimuli) will be provided to users.
- **Questions:** Pseudo-randomize questions will be utilized to provide 2 questions randomly (top 5 largest or top 5 smallest).
- **Visualization Tool:** The same visualization development tool will be utilized to have similar base features.
- **Data Source:** The same data source (dataset) will be utilized to power the visualization.
- **Training:** The same standardized training will be provided to the users.

## Evaluation Criteria:

Descriptive statistics (primarily means) will be used to measure the difference in the data distributions between the control and test visualization. These statistics will be analyzed for time, accuracy, and number of interactions.

Our primary evaluation criteria will be time based on the assumption that if a user is given enough time, they can complete enough clicks and achieve a high level of accuracy.

We expect to see a positive correlation between time and accuracy and a negative correlation between time and number of clicks. A superior visualization will:

- Equal to or more in **Accuracy** of answers
- Equal to or less in **Time** to obtain answers
- Equal to or less in **Number of Interactions**

## Procedure (Methodology)

The visualization evaluation will be instantiated with the following methodology:

1. **Brief Training:** Each user will receive a brief overview of the chart tool and how to interact with key features.
2. **Task 1 Initiate:** The user will be provided with the first task and the first chart to support the task.
  - a. User will review the the provide task question
  - b. User will review and interact with the provided chart (in an html page)
  - c. User will generate an answer to the task and provide it as a solution
3. **Task 1 Data Collection:** The evaluator will collect data on the user (time taken to complete, accuracy of the answer, and clicks made in the visualization) while they answer the task question.
  - a. Record start time and stop time of task
  - b. Grade the answer based on number of correct parts
  - c. Record the number of clicks the user completed
4. **Task 2 Initiate:** The user will be provided with the second task and the second chart to support the task. (see the above substeps for what the user will do)
5. **Task 2 Data Collection:** The evaluator will collect data on the user (time taken to complete, accuracy of the answer, and clicks made in the visualization) while they answer the task question. (see the above substeps for what the user will do)
6. **Closure:** Allow the user to explore the visualizations based on their interests and record any quantitative feedback. Close out the session.
7. **Analysis and Findings:** Analyze the data collected, draw conclusions, and document findings.

## People Recruited

I will **recruit 3 people**: my wife and my 2 in-laws that live in our house. While a small sample of 3, it will give me a couple of repetitions and by utilizing my user with a **within participants factor** where each person sees all conditions.



# Results

Table of Results Collected During the Experiment.

User	Test 1				Test 2			
	Test Setup	Time	Accuracy	Clicks	Test Setup	Time	Accuracy	Clicks
Laura (wife)	Q1 C	32 s	0/3	14 c	Q2 T	23 s	3/3	1 c
Mary (mother in-law)	Q2 T	14 s	3/3	0 c	Q1 C	89 s	3/3	14 c
Scott (father in-law)	Q1 T	72 s	3/3	0 c	Q2 C	210 s	3/3	43

- \* **Question 1 (Q1):** What are the top 3 items at L3 that increase the Consumer Price Index (CPI)?
- \* **Question 2 (Q2):** What are the top 3 items at L3 that decrease the Consumer Price Index (CPI)?
- \* **Control (C):** Complete the question with the Drill-Down Chart
- \* **Test (T):** Complete the question with the Top Items Chart

## Calculated Results for Each Chart Type

Record	Control			Test		
	Time	Accuracy	Clicks	Time	Accuracy	Clicks
1	32 sec	0%	14	23 sec	100%	2
2	89 sec	100%	14	14 sec	100%	3
3	210 sec	100%	43	72 sec	100%	2
Average	110 sec	66%	24	36 sec	100%	2

## Conclusion

Based on the experiment and the results the new visualization is statistically better in all 3 measurements:

- **3x** more efficient with the **time** required to complete the comparison
- **33%** more **accurate** answers to the task's question
- **10x** fewer **interaction to complete** the comparison

Overall, the top 10 visualization bar chart developed for the task is accepted as the superior visualization and should be the base for future iterations and experimentation.

# Findings

The following is a synthesis of findings, including what elements of the approach worked well and what elements should be refined in future iterations.

## What Worked Well

- **Pseudo-Randomization of Tasks and Experimental Stimuli:** It worked well to vary the order of interacting with the test stimuli and the control stimuli. Almost all the participants shared that they felt like the second task was easier after puzzling through the first task.
- **2 Similar but Different Questions for Test:** It was good to have 2 tasks sub questions when doing the within participants factor. This allowed the experiment to have 2 unique answers and not have a participant rely on recall for the prior question. I think the participants also appreciated learning something about the CPI on both ends (what drove it up and what drove it down in February).
- **Narrow Focus for Test Visualization:** The narrow focus of the test visualization provided better accuracy and efficiency. On the qualitative side, it seemed to be less overwhelming for my novice testers given its lack of interaction. It really worked well to build a visualization to purpose as opposed to building a general visualization to answer all questions.

## What Should be Refined

- **Scale of Participants and Automation:** Gathering the evaluation criteria by hand was not very accurate, especially when trying to measure clicks, time, and accuracy of solution. Adding some computer program to assist this testing would be useful. Also, adding some kind of automation tool would support scaling the experiment to more users. For a better sample size, I think I would use mechanical turk to provide the test to a variety of users and increase the sample size by 1 to 2 orders of magnitude.
- **Training on the Tool and How to Use It:** My training on the tool for the users got better as I delivered it, but I underestimated how much background needed to be provided. I think at scale, I would want to do some testing of this training as a thinkaloud to calibrate for the larger test. By tester 3, I think the training delivery had improved and the statistics recorded were also significantly better for that tester.
- **Interactivity on Chart:** In future tests or iterations of the experiment, I'd like to add some interactivity back to the Test chart and allow a user to vary the metric for the top 10 or select the top k values to show. These would be nice value-added features to make the visualization more flexible.

# Appendix A: Part 1: Finding your Data

Locate a dataset that you are interested in working with. The data should be sufficiently complex that you can ask lots of questions about it and engage in creative design techniques, but not so complex that you need specialized hardware or algorithmic approaches to analyze. While you are welcome to use any data you'd like, I recommend that your datasets are tabular (e.g., CSV, TSV, SQL, etc.), contain 5,000 or fewer data points (on the order of one hundred or so tends to be sufficiently interesting without causing lag in Altair), and is data that you're comfortable discussing as part of the course (e.g., avoid data that is overly private or classified).

Discuss your dataset, including the data's source, key attributes/dimensions of the data, and your goals for working with that data (i.e., what are the key questions you want to answer). Identify existing relevant visualizations for working with that data (either using the same data, showing the same concepts, or just that might provide some inspiration) and critique those visualizations based on the practices from this module. What works well? What might need improvement or to change to answer your target questions?

## The Data Source

On a monthly basis, the Bureau of Labor Statistics within the Department of Labor posts a monthly news release on changes to the Consumer Price Index (CPI). An example of the report can be found here: <https://www.bls.gov/news.release/pdf/cpi.pdf>. Along with the news release, they update the data source and a selection of charts to accompany the news release.

The data source can be found here: **CPI-U News Release Companion File:** <https://www.bls.gov/web/cpi/cpi-u.xlsx>

**Raw Versions** of the data and **Attribute Reference Data** can be found here: <https://download.bls.gov/pub/time.series/cu/>

## Key Attribute and Dimensions in the Data:

- **Item Code** - Identifies items for which data observations pertain.
- **Item Name** - Full names of items.
- **Item Code Parent** - How a specific item rolls-up into a larger grouping or category as the CPI is calculated
- **Item Code Level** - The level of the item code hierarchy where the item code is located
- **Period Year** - Identifies the year when the observation occurred
- **Period Month** - Identifies the month, when the period when the data is observed monthly
- **Relative Importance** - the weight associated with the item code (all item codes at a specific level will sum to 100)
- **Unadjusted Percent Change (Twelve Months)** - The percentage change for the specific item code when comparing the current month to the same month 1 year (12 months) ago.
- **Unadjusted Effect on All Items (12 Months)** - The contribution of the specific item to the CPI for All Items when comparing the current month to the same month 1 year (12 months) ago. This can be a positive or negative contribution to the overall CPI. This field, in conjunction with the Relative Importance can be used to compute a CPI for a subset of items.

## Goals for Working with that Data

(i.e., what are the key questions you want to answer).

While the CPI news release and charts are thorough, they focus on the top level CPI number. This is effective at describing the why behind inflation, but given the amount of aggregation, this package falls short of successfully communicating the "so what" of changes in CPI and inflation to the average person.

I'd like to work with the data set to better answer the following key questions:

- What is driving the month over month changes?

- Are there offsetting increases and decreases that are obscured in the highly aggregated metric? How can I make them more transparent?
- How can I enable a citizen to explore the data in an effective way that enables them to discover relevant pricing changes that impact their lives?
- Is it possible to enable a user to shop their own “basket” with the data to customize a CPI index specific to the person to better plan for the month and year ahead?

## Identify Existing Relevant Visualizations

Identify existing relevant visualizations for working with that data (either using the same data, showing the same concepts, or just that might provide some inspiration) and critique those visualizations based on the practices from this module. What works well? What might need improvement or to change to answer your target questions?

### Example 1: Table of Consumer Price Index Statistics

Excerpt from: **CPI-U News Release Companion File:**

<https://www.bls.gov/web/cpi/cpi-u.xlsx>

**Table 1. Consumer Price Index for All Urban Consumers (CPI-U): U.S. city average, by expenditure category, February 2023**  
[1982-84=100, unless otherwise noted]

Expenditure category	Relative importance Jan. 2023	Unadjusted indexes			Unadjusted percent change		Seasonally adjusted percent change		
		Feb. 2022	Jan. 2023	Feb. 2023	Feb. 2022- Feb. 2023	Jan. 2023- Feb. 2023	Nov. 2022- Dec. 2022	Dec. 2022- Jan. 2023	Jan. 2023- Feb. 2023
All items.....	100.000	283.716	299.170	300.840	6.0	0.6	0.1	0.5	0.4
Food.....	13.521	292.794	319.136	320.569	9.5	0.4	0.4	0.5	0.4
Food at home.....	8.727	274.568	301.435	302.483	10.2	0.3	0.5	0.4	0.3
Cereals and bakery products.....	1.169	306.193	349.294	350.772	14.6	0.4	0.6	1.0	0.3
Meats, poultry, fish, and eggs.....	1.846	301.240	322.737	321.803	6.8	-0.3	0.8	0.7	-0.1
Dairy and related products.....	0.813	242.386	272.040	272.271	12.3	0.1	0.2	0.0	0.1
Fruits and vegetables.....	1.508	333.675	351.029	351.249	5.3	0.1	-0.1	-0.5	0.2
Nonalcoholic beverages and beverage materials.....	1.046	192.210	213.359	215.925	12.3	1.2	0.5	0.4	1.0
Other food at home.....	2.344	237.143	264.746	266.612	12.4	0.7	0.7	0.7	0.3
Food away from home <sup>1</sup> .....	4.795	320.880	345.677	347.869	8.4	0.6	0.4	0.6	0.6
Energy.....	7.076	267.771	283.330	281.673	5.2	-0.6	-3.1	2.0	-0.6
Energy commodities.....	3.556	313.522	305.843	309.280	-1.4	1.2	-7.2	1.9	0.5
Fuel oil <sup>1</sup> .....	0.147	384.179	455.595	419.629	9.2	-7.9	-16.6	-1.2	-7.9
Motor fuel.....	3.344	307.422	297.413	302.153	-1.7	1.6	-6.9	2.3	0.9
Gasoline (all types).....	3.246	305.959	294.759	299.774	-2.0	1.7	-7.0	2.4	1.0
Energy services.....	3.520	235.017	272.840	266.353	13.3	-2.4	1.9	2.1	-1.7
Electricity.....	2.579	236.453	266.528	266.887	12.9	0.1	1.3	0.5	0.5
Utility (piped) gas service.....	0.941	226.491	285.407	258.976	14.3	-9.3	3.5	6.7	-8.0
All items less food and energy.....	79.403	288.059	301.962	304.011	5.5	0.7	0.4	0.4	0.5
Commodities less food and energy.....	21.325	164.559	165.340	166.246	1.0	0.5	-0.1	0.1	0.0
Apparel.....	2.525	127.868	127.875	132.039	3.3	3.3	0.2	0.8	0.8
New vehicles.....	4.299	168.027	177.276	177.817	5.8	0.3	0.6	0.2	0.2
Used cars and trucks.....	2.605	212.040	185.857	183.241	-13.6	-1.4	-2.0	-1.9	-2.8
Medical care commodities <sup>1</sup> .....	1.459	384.304	395.981	396.560	3.2	0.1	0.1	1.1	0.1
Alcoholic beverages.....	0.843	269.128	282.286	282.377	4.9	0.0	0.7	0.4	-0.3
Tobacco and smoking products <sup>1</sup> .....	0.493	1,314.838	1,388.790	1,402.834	6.7	1.0	-0.1	0.7	1.0
Services less energy services.....	58.078	363.672	387.258	390.070	7.3	0.7	0.6	0.5	0.6
Shelter.....	34.393	344.607	369.585	372.510	8.1	0.8	0.8	0.7	0.8
Rent of primary residence.....	7.521	359.627	388.372	391.141	8.8	0.7	0.8	0.7	0.8
Owners' equivalent rent of residences <sup>2</sup> .....	25.385	353.512	379.328	381.844	8.0	0.7	0.8	0.7	0.7
Medical care services.....	6.591	586.177	601.551	598.778	2.1	-0.5	0.3	-0.7	-0.7
Physicians' services <sup>1</sup> .....	1.838	407.984	415.197	413.034	1.2	-0.5	0.1	-0.1	-0.5
Hospital services <sup>1,3</sup> .....	1.934	371.745	385.064	384.996	3.6	0.0	1.3	0.5	0.0
Transportation services.....	5.752	333.919	376.743	382.801	14.6	1.6	0.6	0.9	1.1
Motor vehicle maintenance and repair <sup>1</sup> .....	1.110	331.130	371.780	372.572	12.5	0.2	1.0	1.3	0.2
Motor vehicle insurance.....	2.528	584.317	658.513	668.806	14.5	1.6	0.7	1.4	0.9
Airline fares.....	0.574	222.227	264.629	281.216	26.5	6.3	-2.1	-2.1	6.4

<sup>1</sup> Not seasonally adjusted.  
<sup>2</sup> Indexes on a December 1982=100 base.  
<sup>3</sup> Indexes on a December 1996=100 base.

### What works well:

This view is nice because it offers a comprehensive view of all statistics calculated for CPI for all product combinations. The indented format allows a user to skim the table and find the associated statistics for a specific item or aggregation in the CPI. For the purpose of finding a specific data point, the vertical nature of the table and the indentation structure allows the user to efficiently scan the items and identify a specific item when searching.

### What might need improvement:

There is no interaction for the view. If you come to it with some knowledge of the data, you are still forced to sift through all the information. A filter to search for a specific value based on text or a filter to limit the granularity of the view, would greatly facilitate a user looking for a specific item in the CPI or review a specific grain of the CPI calculation. The ability to

sort the view would help the user identify a specific point that meets a criteria, like largest change. Using a diverging color scale, we can help the user identify the magnitude of change more easily.

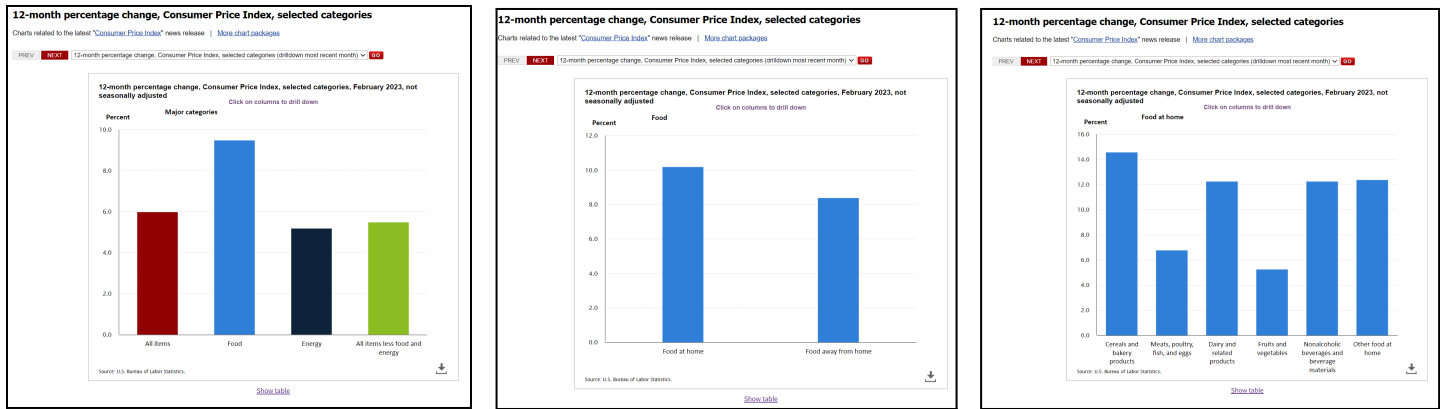
**What change is needed to answer your target questions:**

I think tables may be a good item to use when we drill to details at lower levels (more granular). In this table, I would like to add the ability to filter on specific attributes or derived attributes. I'd also like to add some coloration to the metrics. Finally, we could consider adding a sparkline to a more narrow view to also visually understand the trend of data for each item in the CPI.

**Example 2: Select Category Bar Chart for 1 Month Percentage Change**

Excerpt from: **12-month percentage change, Consumer Price Index, selected categories**

<https://www.bls.gov/charts/consumer-price-index/consumer-price-index-by-category-line-chart.htm>



**What works well:**

The chart uses categorical color schemes to separate the different top level groups of the CPI. The user can also click to drill in on a specific category to see the CPI for each more granular category. The use of bars is useful for comparative analysis and is a good selection. Labels are in alphabetical order, which is useful for finding the data point in the same section, based on repeated use.

**What might need improvement:**

After each click, the parent chart disappears. While the hierarchy isn't super deep, it does lead to some level of getting lost in the "data fog" and you become disoriented about what level you are viewing, what the parent group was, and what you have already clicked into. This can be improved by keeping the parent charts in the view for context (in this case), it might work best to click to expand a category (see its children) and keep the other bars in the view. Alternatively, it could work well to just have 3 panes that are connected. One final option could be to include a constant marker line to represent the average of the parent and label the parent group on the line to give the better context.

**What change is needed to answer your target questions:**

I'd like to enable this view with click to expand to allow the user to select bars. If I want a user to be able to shop items in the overall CPI. I could use a bar chart like this and dynamically update the overall CPI based on their selection of metrics. That could be a really powerful tool for interaction.

**Example 3: Select Category Bar Chart for 1 Month Percentage Change**

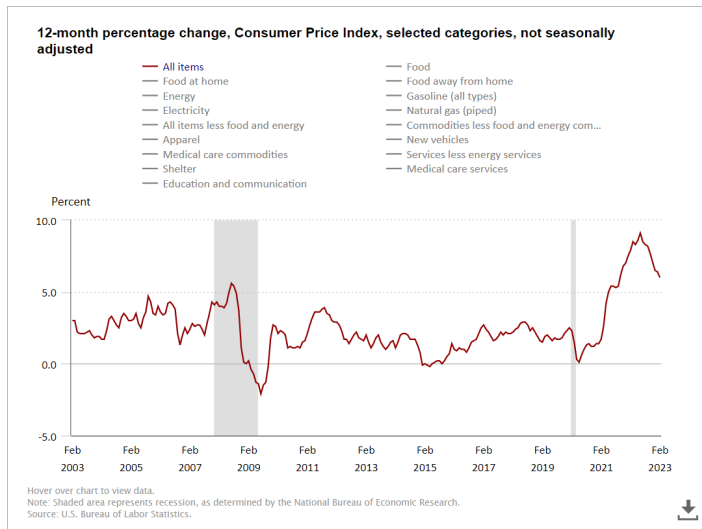
Excerpt from: **12-month percentage change, Consumer Price Index, selected categories**

<https://www.bls.gov/charts/consumer-price-index/consumer-price-index-by-category-line-chart.htm>

## 12-month percentage change, Consumer Price Index, selected categories

Charts related to the latest "Consumer Price Index" news release | [More chart packages](#)

[PREV](#) [NEXT](#) 12-month percentage change, Consumer Price Index, selected categories (past 20 years) [GO](#)



### What works well:

We can see a large trend of percentage change in CPI (aka inflation or deflation). This gives us a lot of context of where it has been. The different categories are also colored appropriately to represent their independent nature. There is interaction to click to add or remove lines.

### What might need improvement:

When you want to see more than 3 lines, the line chart begins to get very cluttered. If we really want to show changes for greater than 10 items, we should consider using a different type of chart that better visualizes a more specific element. There is a lot of history in the chart. Allowing a user to zoom could reduce the noise and make it easier to focus on the more recent trend. We know there is a target inflation number that hasn't changed. It might be more useful to focus on the how above or below this number we are or add a reference line on the chart.

### What change is needed to answer your target questions:

I'd like to show this chart as a grouping of smaller sparkline charts with some key statistics. I'd also like the user to be able to zoom in with a slider or an interaction with a timeline chart. I'd like to add a reference line or show the delta between the target inflation number with gradient coloration of the bars to indicate how far from the target the numbers are. This would need to be based on absolute value, b/c being too low is just as bad as being too high.

## Other Charts Not Included in Writeup

### Example A: CPI-U News Release Companion File

see: <https://www.bls.gov/web/cpi/cpi-u.xlsx>

Chart 1. One-month percent change in CPI for All Urban Consumers (CPI-U), seasonally adjusted, Feb. 2022 - Feb. 2023

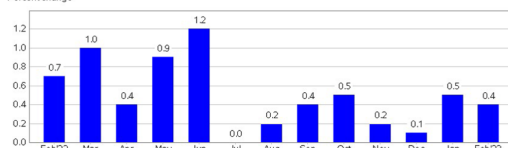


Chart 2. 12-month percent change in CPI for All Urban Consumers (CPI-U), not seasonally adjusted, Feb. 2022 - Feb. 2023

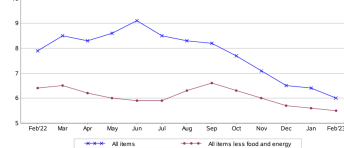


Table A. Percent changes in CPI for All Urban Consumers (CPI-U): U.S. city average

	Seasonally adjusted changes from preceding month							Un-adjusted 12-mos. ended Feb. 2023
	Aug. 2022	Sep. 2022	Oct. 2022	Nov. 2022	Dec. 2022	Jan. 2023	Feb. 2023	
All items.....	0.2	0.4	0.5	0.2	0.1	0.5	0.4	6.0
Food.....	0.8	0.8	0.7	0.6	0.4	0.5	0.4	9.5
Food at home.....	0.8	0.7	0.5	0.6	0.5	0.4	0.3	10.2
Food away from home¹.....	0.9	0.9	0.9	0.5	0.4	0.6	0.6	8.4
Energy.....	-3.9	-1.7	1.7	-1.4	-3.1	2.0	-0.6	5.2
Energy commodities.....	-8.0	-4.1	3.7	-2.1	-7.2	1.9	0.5	-1.4
Gasoline (all types).....	-8.4	-4.2	3.4	-2.3	-7.0	2.4	1.0	-2.0
Fuel oil¹.....	-5.9	-2.7	19.8	1.7	-16.6	-1.2	-7.9	9.2
Energy services.....	1.8	1.2	-0.7	-0.6	1.9	2.1	-1.7	13.3
Electricity.....	1.2	0.8	0.5	0.5	1.3	0.5	0.5	12.9
Utility (piped) gas service.....	3.5	2.2	-3.7	-3.4	3.5	6.7	-8.0	14.3
All items less food and energy.....	0.6	0.6	0.3	0.3	0.4	0.4	0.5	5.5
Commodities less food and energy commodities.....	0.4	0.0	-0.1	-0.2	-0.1	0.1	0.0	1.0
New vehicles.....	0.8	0.7	0.6	0.5	0.6	0.2	0.2	5.8
Used cars and trucks.....	-0.2	-1.1	-1.7	-2.0	-2.0	-1.9	-2.8	-13.6
Apparel.....	0.3	0.0	-0.2	0.1	0.2	0.8	0.8	3.3
Medical care commodities¹.....	0.2	-0.1	0.0	0.2	0.1	1.1	0.1	3.2
Services less energy services.....	0.6	0.8	0.5	0.5	0.6	0.5	0.6	7.3
Shelter.....	0.7	0.7	0.7	0.6	0.8	0.7	0.8	8.1
Transportation services.....	1.0	1.9	0.6	0.3	0.6	0.9	1.1	14.6
Medical care services.....	0.7	0.8	-0.4	-0.5	0.3	-0.7	-0.7	2.1

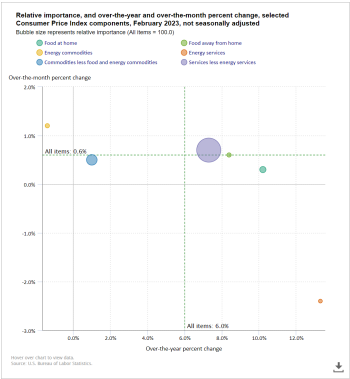
Example B: Relative importance, and over-the-year and over-the-month percent change

See: <https://www.bls.gov/charts/consumer-price-index/consumer-price-index-relative-importance.htm>

Relative importance, and over-the-year and over-the-month percent change

Charts related to the latest "Consumer Price Index" news release | [More chart packages](#)

Previous NEXT Relative importance, and over-the-year and over-the-month percent change (bubble chart) 1 of 2



# Appendix B: Part 2: Sketching your Data

Your Module 1 discussion post identified some high-level goals for working with a dataset of interest to you. In this post, you will expand on those goals to characterize your target problem and develop some low-fidelity prototypes for working with that data. First, identify two to three tasks you would wish to complete with your data, identifying:

- Why is a task pursued? (goal)
- How is a task conducted? (means)
- What does a task seek to learn about the data? (characteristics)
- Where does the task operate? (target data)
- When is the task performed? (workflow)
- Who is executing the task? (roles)

Then, sketch a set of preliminary low-fidelity prototypes for addressing these tasks with the given data. You may either sketch freeform or use the Five Design Sheets approach to generate these prototypes (hand-sketched on paper is fine). Upload a copy of your sketches as part of your post.

## Recap of Goals Initially Identified

As a review, in part 1, I identified the following goals to analyze and understand when working with the Consumer Price Index Data and developing better visualizations.

- What is driving the month over month changes?
- Are there offsetting increases and decreases that are obscured in the highly aggregated metric? How can I make them more transparent?
- How can I enable a citizen to explore the data in an effective way that enables them to discover relevant pricing changes that impact their lives?
- Is it possible to enable a user to shop their own “basket” with the data to customize a CPI index specific to the person to better plan for the month and year ahead?

## Task 1: Citizens Analyzing CPI for Understanding and Relevance

### **Why is a task pursued? (goal)**

Explore data to assess relevant items to my personal life and identify how they are changing.

### **How is a task conducted? (means)**

Navigate between aggregation levels of CPI to discover patterns and fluctuations (increasing and decreasing) for items that makeup the Consumer Price Index metric.

### **What does a task seek to learn about the data? (characteristics)**

Use outcomes to make purchasing decisions and possibly help inflation by making smart buying decisions.

### **Where does the task operate? (target data)**

Operate with the target data of Consumer Price Index dataset from the Bureau of Labor Statistics.

### **When is the task performed? (workflow)**

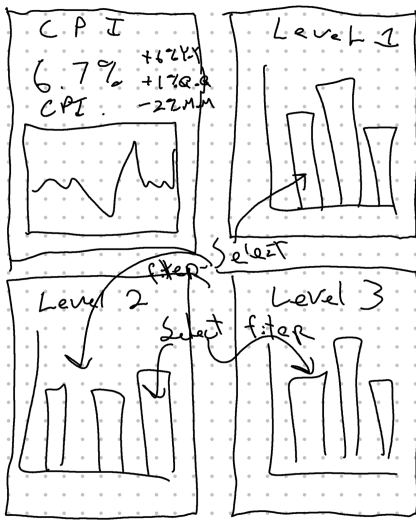
Perform monthly to understand changes in pricing and make better household purchasing decisions.

### **Who is executing the task? (roles)**

United States consumers and citizens.



## Sketch a Set of Preliminary Low-fidelity Prototypes



## Task 2: Policy Makes Analyzing CPI Influencers

### Why is a task pursued? (goal)

Understand what is primarily driving CPI changes (inflation) in the current period.

### How is a task conducted? (means)

Review what the top x influencers are to CPI in the positive and negative direction.

### What does a task seek to learn about the data? (characteristics)

Use outcomes to understand what is driving the CPI changes (winning and losing) and where to focus efforts to combat inflation.

### Where does the task operate? (target data)

Operate with the target data of Consumer Price Index dataset from the Bureau of Labor Statistics.

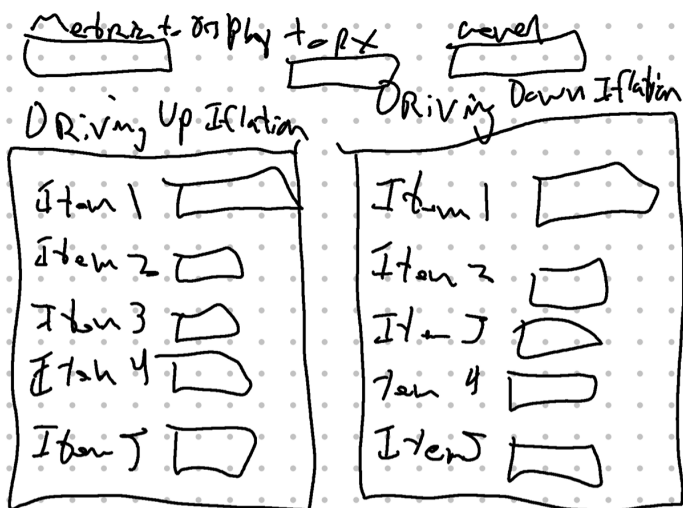
### When is the task performed? (workflow)

Perform monthly to understand changes in pricing and make better policy choices to focus effort.

### Who is executing the task? (roles)

Economists or policy making that can influence prices or consumer behaviors.

## Sketch a Set of Preliminary Low-fidelity Prototypes



## Task 3: CPI Changes MoM

### Why is a task pursued? (goal)

Understand how the CPI is fluctuating month over month and any offsetting influencers to the aggregate metrics.

### How is a task conducted? (means)

Review charts and identify positive and negative pressures on CPI month over month

### What does a task seek to learn about the data? (characteristics)

Use outcomes to make purchasing decisions and possibly help inflation by making smart buying decisions.

### Where does the task operate? (target data)

Operate with the target data of Consumer Price Index dataset from the Bureau of Labor Statistics.

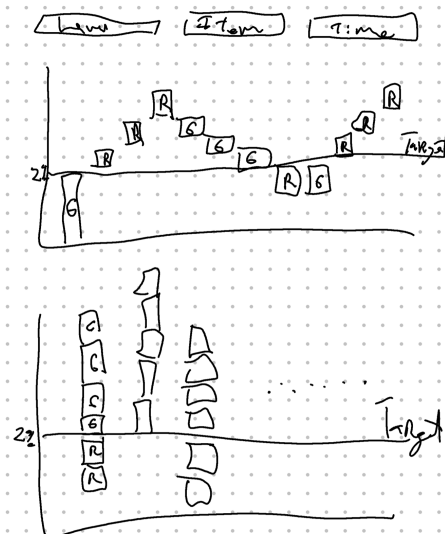
### When is the task performed? (workflow)

Perform monthly to understand hidden trends in the CPI.

### Who is executing the task? (roles)

United States consumers and citizens.

### Sketch a Set of Preliminary Low-fidelity Prototypes



## Appendix C: Part 3: A Plan for Evaluation

In your previous post, you identified a series of tasks and goals for your visualization as well as some preliminary design ideas. We'll jump ahead a few steps and start to think about how we might evaluate our design approach. Outline a preliminary evaluation that addresses your core goals with the visualization. Make sure your evaluation discusses:

- The target question you want to answer
- The people you would recruit to answer that question
- The kinds of measures you would use to answer your data (e.g., insight depth, use cases, accuracy) and what these measures would tell you about the core question
- The approach you will use to answer that question (e.g., a journaling study, a formal experiment, etc.)
- How would you instantiate those methods (i.e., what would your participants do?)
- What criteria would you use to indicate that your visualization was successful
- The target question you want to answer

### Target Question

*The target question you want to answer.*

**Question:** What CPI representation lets people most accurately determine the top X and bottom X influencers to CPI at a given level of items.

**Hypotheses:** A sorted top x and bottom x plot will be most accurate.

## Recruited People

*The people you would recruit to answer that question*

I will recruit my wife and my 2 inlaws that live in our house. While a small sample of 3, it will give me a couple of repetitions.

If I wanted to scale this evaluation, I would use mechanical turk to increase my participants and increase my sample size.

## Measures and Outcomes

*The kinds of measures you would use to answer your data (e.g., insight depth, use cases, accuracy) and what these measures would tell you about the core question*

We would look at 3 key measures to determine the quality of the solution to answer our core questions:

- **Time:** How long does it take the user to complete the given task? (lower is better)
- **Accuracy:** How Accurate can people complete a given task? (higher is better)
- **Number of Interactions:** How many steps does it take to complete the given task? (lower is better)

## Approach

*The approach you will use to answer that question (e.g., a journaling study, a formal experiment, etc.)*

For the evaluation, I will use a Summative Experimental Evaluation as a Quantitative Evaluation. I will use a within participants factor where each person sees all conditions. Given the medium (likely a local copy) I will be present to manually measure the dependent variables and collect data. I will also be able to collect any qualitative data from the users from Thinkalouds that follow the experimental task.

### Experimental Tasks:

[Comparison] Determine the top 3 influential segments for CPI (either direction):

- Determine the top 3 segments increasing CPI for the month
- Determine the top 3 segments decreasing CPI for the month

### Experimental Stimuli:

- **Control:** Recreation of the BLS Bar Chart of 12-month percentage change, Consumer Price Index
- **Test:** New bar chart layout focused on top X influencers to increase CPI and top X influencers to decrease CPI

### Independent Variables:

Configuration of Bar Chart to explore and display dataset (layout, filtering, and features).

### Dependent Variables:

- **Time:** How long does it take the user to complete the given task?
- **Accuracy:** How Accurate can people complete a given task?
- **Number of Interactions:** How many steps does it take to complete the given task?

#### Control Variables:

- **Experimental Stimuli:** Pseudo-randomize order of visualizations (experimental stimuli) will be provided to users.
- **Questions:** Pseudo-randomize questions will be utilized to provide 2 questions randomly (top 5 largest or top 5 smallest).
- **Visualization Tool:** The same visualization development tool will be utilized to have similar base features.
- **Data Source:** The same data source (dataset) will be utilized to power the visualization.
- **Training:** The same standardized training will be provided to the users.

## Methodology

*How would you instantiate those methods (i.e., what would your participants do?)*

8. **Brief Training:** Each user will receive a brief overview of the chart tool and how to interact with key features.
9. **Task 1 Initiate:** The user will be provided with the first task and the first chart to support the task.
  - a. User will review the the provide task question
  - b. User will review and interact with the provided chart (in an html page)
  - c. User will generate an answer to the task and provide it as a solution
10. **Task 1 Data Collection:** The evaluator will collect data on the user (time taken to complete, accuracy of the answer, and clicks made in the visualization) while they answer the task question.
  - a. Record start time and stop time of task
  - b. Grade the answer based on number of correct parts
  - c. Record the number of clicks the user completed
11. **Task 2 Initiate:** The user will be provided with the second task and the second chart to support the task.  
(see the above substeps for what the user will do)
12. **Task 2 Data Collection:** The evaluator will collect data on the user (time taken to complete, accuracy of the answer, and clicks made in the visualization) while they answer the task question.  
(see the above substeps for what the user will do)
13. **Thinkalouds + Closure:** Collect any qualitative data and feedback from the user regarding the charts and their usefulness. Allow the user to explore anything else in the visualizations based on their interests. Close out the session.

## Evaluation Criteria

*What criteria would you use to indicate that your visualization was successful*

Descriptive statistics (primarily means) will be used to measure the difference in the data distributions between the control and test visualization. These statistics will be analyzed for time, accuracy, and number of interactions.

Our primary evaluation criteria will be time based on the assumption that if a user is given enough time, they can complete enough clicks and achieve a high level of accuracy.

We expect to see a positive correlation between time and accuracy and a negative correlation between time and number of clicks. A superior visualization will:

- Equal to or more in **Accuracy** of answers
- Equal to or less in **Time** to obtain answers
- Equal to or less in **Number of Interactions**

# Appendix D: Part 4: Final Project

Throughout the Modules, you have found a dataset, characterized the corresponding goals and tasks you want to conduct with that data, designed preliminary approaches, and outlined how you would evaluate those approaches. For your final project, you will put these ideas into practice by executing on the project plan outlined in your prior posts.

For this project, you will implement a visualization using your data from Module 1 and preliminary low-fidelity prototypes from Module 2 to address your stated goals. You may implement this visualization using either Altair or another platform of your choice. Once implemented, conduct your evaluation based on the plan outlined in your Module 3 discussion post, making sure to conduct your evaluation with at least three people. You may refine any of your prior plan to reflect your evolving understanding of the challenges you are addressing. Be sure to address how your plan has changed from these earlier posts as part of your discussion.

Your final project post should include:

- A brief recap of your data, goals, and tasks, focusing on those that most directly influence your design
- Screenshots of and/or a link to your visualization implementation (see below for additional guidance)
- A summary of the key elements of your design and accompanying justification
- A discussion of your final evaluation approach, including the procedure, people recruited, and results. Note that, due to the difficulty of recruiting experts, you can use colleagues, friends, classmates, or family to evaluate your designs if experts or others from your target population are unavailable.
- A synthesis of your findings, including what elements of your approach worked well and what elements you would refine in future iterations.

Guidance and platforms for deploying Altair visualizations online include:

- <https://matthewkudija.com/blog/2018/06/22/altair-interactive/>
- <https://towardsdatascience.com/add-animated-charts-to-your-dashboards-with-streamlit-python-f41863f1ef7c>
- <https://towardsdatascience.com/creating-interactive-jupyter-notebooks-and-deployment-on-heroku-using-voila-aa1c115981ca>

## Develop Solution

- Build solution in jupyter notebooks in colab.
- Publish visualizations to files and host on google drive
- **Link to Draft Data Source:** [CPI-U](#)
- **Link to Final Data Source:** [CPI-U - Output.tsv](#)
- **Link to Jupyter Notebook:** [CPI Visualization Project.ipynb](#)
- **Link to Control File for Visualization:** [CONTROL - L3 final combo.html](#)
- **Link to Test File for Visualization:** [TEST - L3 top final combo.html](#)

## Recap

A brief recap of your data, goals, and tasks, focusing on those that most directly influence your design

See main body of the report.