World Council for City Data
(in partnership with Focus21)

Joel Schellenberger, Cameron Fong, Kathy Nguyen, Mikayla Ferraro, Nicole Chow, Tyler Oakes, Cedric Zheng

April 2019

Over the last seven weeks, our team has been working diligently to create a solution for the concerns presented to us by our client, Focus21 and the World Council on City Data (WCCD).

Initially our client presented us their concern of 'scalability' for managing and presenting the large data sets provided by the WCCD. Through diligent research and user validation we came to the conclusion that we would focus on the redesign of the website to ensure more effective usability by simplifying the appearance of the data to make it more engaging to work with, while removing certain barriers of entry from the current offering.

We began by conducting research through literature regarding best practices when designing sites for data visualization with large data sets. We also completed competitive analysis' in order to analyze other well-designed websites with a focus on data visualization to capture the strengths and weaknesses of other websites of this nature. Following this, we conducted ideation workshops to generate as many ideas as we could. We conducted phases of individual and team idea generating, which was followed by heat mapping and selecting best elements of multiple designs to help guide us to one idea to be prototyped.

This led us into designing new and refined features that give the user access to the types of data they'd find valuable in a visual comparison, allowed them to access it in a way that was simple enough to comprehend on first viewing, and empowered them to understand where the data was coming from in relation to ISO 37120.

During our usability testing, we found some common and informative concerns that helped us shape the final design. Through our findings, we have been able to work through a redesign of the WCCD website that is visually appealing, usable, and scalable, as well as develop recommendations for the future of the site.

Contents

Long-term project goal:

Create high-fidelity mockups, based on user research and validated by testing, of an effective, scalable data visualization interface for the WCCD website.

UNDERSTAND

Week One
On/Before February 8th

Understand our problem space:

- Evaluate current site offering
- Gather resource requirements
- Gather context from client
- Seek out other large scale data visualization projects

Project Plan

- Detailed roadmap
- Clearly define goals / expectations / deliverables

UNDERSTAND

Week Two
On/Before February 15th

"Competitive" Analysis

- Evaluate other large scale data visualization projects based on interfaces, target users of "competitors," what works and what doesn't work for those projects
- Work to understand design decisions when displaying large amounts of data (considerations including density, default display settings, scalability, and more)

Usability Testing Framework

- Establish usability tests, pertaining to large-scale data visualization projects
- Base testing on a number of factors, including data gathered from "competitive" analysis, heuristics, etc.
- Work to understand the behaviors users expect of sites with large-scale, dense data at their core

UNDERSTAND

Week Three
On/Before March 1st

Continued External Research

- Review literature and documentation on large-scale data visualization projects with the goal of developing a deep understanding of patterns, trends, assumptions and goals for projects in a similar space
- Bring together research findings into comprehensive reference document that helps communicate key findings in an easily-digestible format for presentation to stakeholders

IDEATE

Week Four
On/Before March 8th

Idea Generation
- Four Step Sketches (Notes, Ideas, Crazy 8s, Solutions)
- Other ideation activities based on findings from research

Review Ideation
- Seek out strongest solutions through activities such as heat mapping, speed critiques, straw polls, etc.
- Establish scope of prototyping activities and which solutions are worth moving forward on
- Utilize storyboarding to plan prototyping phase out more clearly and help iterate on the previously carried out user testing process / interview frame

PROTOTYPE

Week Five
On/Before March 15th

Prototyping
- Choose tools which provide a strong balance of fidelity, speed and functionality
- Work in individual roles to build out prototypes that communicate our ideas / solutions on a reasonable scale re: time constraints

Preparations for Testing
- Stitch together individual efforts into cohesive prototype(s) and correct any mistakes before testing
- Adjust interview framework for new prototypes
- Identify tools to use during testing phase that would be helpful to track more thorough data
- Run through prototypes, making necessary adjustments

TEST

Week Six
On/Before March 22nd

Testing / Validation
- Conduct interviews / contextual inquiries utilizing tools established during testing preparations (webcam, screen recording, etc.)

Reflection
- Make notes on the testing results, attempting to understand what works and what doesn't work for the users and how improvements can be made
- Compare notes and understandings of findings to ensure coherence going into presentation phase

REFLECT

Week Seven

On / Before March 29th

Prototype Refinements

- Work in a team-based context to refine our prototypes based on data collected during our week of testing
- Make necessary adjustments to move project forward
- Prepare recommendations for actions which may be out of reach due to time constraints, but provide direction based on testing
-

Presentation Prep

- Prepare prototype demo for larger scale presentation
- Prepare slide deck, reflecting on design process and findings from user testing to present learnings
- Prepare questions to ask audience to gather other feedback points on project and solution

NOCturne Display

Jordan Riley Benson &
Rajiv Ramarajan

https://doi-org.libproxy.wlu.ca/10.1145/2609876.2609886

Key Points

- Focused on the story being told
- Understood the limits of the medium
- Utilized grouping and ranking
- Designed a non-interactive overview to test assumptions

- NOCturne display was created for the second VAST challenge in 2013; aiming to connect a network of a million or more devices, with a focus on the improving scalability within large and growing networks.

- An initial challenge was presented when considering what the average user could see; understanding this consideration would help determine what sort of display size was needed for large-scale visualization. This estimate not only helped to determine the size of the display needed but also helped the team visualize solutions that did not include shrinking icons and text just to be able to fit more on the screen.

- NOCturne looks to turn data into graphs, displaying data of the past and making predictions of future trends. In order to navigate the complexities of the problem space the team limited the amount of data being shown on a given graph to 4 tracked metrics -- these metrics could be changed out based on what the individual's business needs are, while enabling users to track small changes over time.

- One large scalability problem that this team faced was representing information on millions of nodes in a way that it is meaningful and effective for the user. Their solution to this problem involved putting all the information into chart form, utilizing chunking in order to provide visual space for the user with tables consisting of icons that reduced clutter on the screen while providing context as to which aspect of the network it was correlated to. The team at NOCturne used hierarchical clustering to determine what elements of the table would be most important to the users business needs, thus displaying them first. In order to show connectivity between the networks NOCturne uses an adjacency matrix, which allows the readability of the data to remain consistent compared to other network visualizations.

- NOCturne featured a world map that could not be zoomed in, causing a visibility issue when displaying information on the map. To combat this, NOCturne increased the size of the map to only display the information that would be essential for a geographic display, limiting the amount that is put on the map in order to increase visibility and readability while also complimenting the other aspects of the project.

How to Design Big Data UX for the Era of Citizen Analysts

Sophia Brooke (2018)

https://uxplanet.org/how-to-design-big-data-ux-for-the-era-of-citizen-analysts-98625db0750d

Key Points

- End users should find the experience effortless, allowing them to interact with data in ways that they find intuitive to navigate

Define Your Users and Their Needs

- Avoid overloading user dashboard with loads of data which is a common mistake when designers fixate on demonstrating the richness of a solutions analytic capabilities
- Many users know there are enormous amounts of data behind charts and graphs but majority of them don't want to see more than necessary to meet their objectives

Design for Less Analytically Adept

- Not all companies who may use these sites will have trained analysts to read the data, so it must be designed for both the novice and expert
- The "winning" UI will be one that can be customized to suit personal preferences-ideally by the user themselves

Essential Considerations While Designing

- Which should be the key metrics visualized to aid decision-making?
- How will different stakeholders be using the data?
- What screen sizes will the data be primarily displayed on?
- Do users need to monitor data in real-time?
- To what extent will users need to interact with the data?
- Don't make assumptions... watch users work.

Use a Prioritization Structure for Dashboards

- Use tabs to separate different reporting elements
- Order those tabs in a structured way from left to right
- Prioritize important data in each tab by highlighting it

Don't Create a Variety Show

- Ease of interpretation should be top prioritity
- Single colour gradients are effective in aiding comprehension

Create Levels of Usability

- Necessity over creative ego, while striving for consistency
- If different business discipline will use it, the UI needs to be complex, the best approach is to have two or three levels of usability (eg. non-technical users will likely want straightforward, easy-to-digest dashboard visualization)
- Management and leadership are likely to need access to menu-based reports and table selection
- Analysts and data scientists might need a combination of menus and keyboard commands, providing access to advanced features

Data Visualization Techniques: From Basic to Big Data...

SAS Institute Inc. (2018)

https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/data-visualization-techniques-106006.pdfalysts-98625db0750d

Key Points

- Understand the data you are trying to visualize and what kind of information you want to communicate
- Determine what you are trying to visualize / what kind of information you want to communicate
- Know your audience; understand how they process visual representation of information
- Use visuals that convey information in the best and simplest form

Basic Charts

- Line Charts
  - Show the relationship of one variable to another
  - Most often used to track changes or trends over time, both individually and when compared to other data
- Bar Charts
  - Used to compare quantities of different categories or groups
  - Can be configured with either vertical or horizontal bars, with lengths and widths representing values
- Scatter Plots (or X-Y Plot)
  - Two-dimensional plot showing joint variation of two data items
  - Can help gain a sense of how spread out the data might be or how closely related the data points are.
  - Can also quickly identify patterns present in data distribution
- Bubble Plots
  - Scatter plot variation where markers are replaced with bubbles
  - Specially useful for data sets with many values or when the values differ by several orders of magnitude
- Pie & Donut Charts
  - Most effective with limited components and when text and percentages are included to describe the content

Handling Large Data Volumes

- Data volume can become an issue -- traditional architectures and software may not be able to process huge amounts of data in a timely manner, thus requiring you to make compromises and aggregate the details you want to visualize
- One potential solution is to use binning (the grouping together of data) on both axes so that you can effectively visualize the big data
- Box plots are another example of how the volume of data can affect the visual being shown. A box plot is a graphical display of five statistics (the minimum, lower quartile, median, upper quartile and maximum) that summarize the distribution of a set of data
- A binned box plot with whiskers enables you to view the distribution of large data and easily see outliers
- A correlation matrix combines big data and fast response times to quickly identify which variables among the millions or billions are related. It also shows how strong the relationship is between variables

Analytical Review of Data Visualization Methods...

Evgeniy Yur'evich Gorodov &
Vasiliy Vasil'evich Gubarev

https://www.researchgate.net/publication/275468420_Analytical_Review_of_Data_Visualization_Methods_in_Application_to_Big_Data

Key Points

- Performance requirements, types of users and approaches to dynamic change are all key to creating an effective solution

Problems With Visualizing Big Data

- Visual noise
    - When the objects in the data are too close to each other, it can lead to a visibility issue in which the user cannot divide the objects to recognize that they are separate.
- Large Image Perception
    - When the user loses the ability to acquire any useful information since all visualization methods are limited by device resolution, aspect ratio, and physical perception limits.
- Information Loss
    - Since one cannot include all data into a visualization technique, the analyst must decide which information to include and exclude. Thus leading to data aggregation and filtration, which can lead to the rejection of what may be interest hidden objects. This filtering to achieve the appropriate information can also consume large amounts of time and resources.
- High Performance Requirements
    - High performance visualizations can often lead to an increase in computing resources (often to refresh or reload the page), this leads to slow reaction and loading times.
- High Rate of Image Change
    - The user may not be able to react to the number of data changes or its intensity on display.

Approaches to Avoid These Problems

- Analyst able to use more than one data representation view at once
    - The researcher must be able to view the data in different charts, graphs and other visualizations to conclude which would be the simplest and most comprehensible to display the data on.
- Active interaction between user and analyzable view
    - The user must be able to zoom in and out of the visualization as well as filter the necessary data with ease without distorting the perception of the whole diagram.
- Dynamic change of factors during working process with view (filters)
    - Selection inversion must be considered. For example, if the user presses on one radio button in a group of radio buttons, resulting in the other buttons losing user selection. This prevents users and analysts from making mistakes when filtering objects or inputting data.

Tufte's Principles & Guidelines of Information Design
Haig Armen

http://courses.haigarmen.com/intd319/files/2017/01/InfoDesign-haig.pdf

Envisioning Information
Edward R. Tufte

https://wikileaks.org/gifiles/attach/158/158881_Tufte%20-%20EnvisioningInformation.pdf

Key Points

- Clutter/confusion are failures of design and not of complexity
- Graphics is intelligence made visible
- Density of design can be a valuable feature when done correctly

1. Comparisons
2. Causality
3. Multivariate Analysis
4. Integration of Evidence
5. Documentation
6. 'Smallest Effective Differences'

Philosophy of Approach

- Assume audience is intelligent
- Don't limit people by "dumbing" the data, allow people to use their abilities to get the most out of it
- Human eye/brain can select / filter / edit / group / structure / highlight / focus / blend / outline / cluster / itemize / sort / abstract / smooth / isolate / idealize / summarize / etc.
- Clutter/confusion are failures of design and not of complexity

Things to Watch Out For

- Compare different time periods
- Use area/volume instead of linear scales to exaggerate differences
- Make use of design variation to exaggerate data variation
- Exaggerate the vertical scale or don't begin at the zero point
- Show only parts of a cycle so that data from other parts of the cycle cannot be used for proper comparison

High Quality Graphics

- Comparative, multivariate, high density, where nearly all the ink is actual data ink... emphasize horizontal direction...
- Exclude bi-lateral (not a geometrical shape) symmetry when redundant
    - Studies show that we often concentrate on one side of a symmetrical figure and only glance at the other side

9 Ways to Make Big Data Visual
import.io

https://www.import.io/post/9-ways-make-big-data-visual/

Key Points

- Only takes one tenth of a second to comprehend a visual scene
- You attach meaning to a symbol in 250 milliseconds
- A picture speaks 1000 words

Heat Map
- Can only really be used with regions that are connected
- Very effective at representing data

Bar Graphs
- Simple, but that is the point
- Can be single, grouped, or clustered

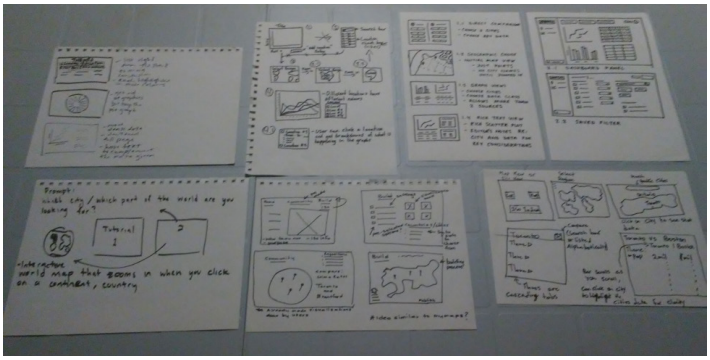
Line Graphs
- Usually used to show change over time
- Use different colours to distinguish different elements

Scatter Plot Graph
- Like a line graph without the line
- Used to show correlation between two things that are different (i.e. flower height and pedal length

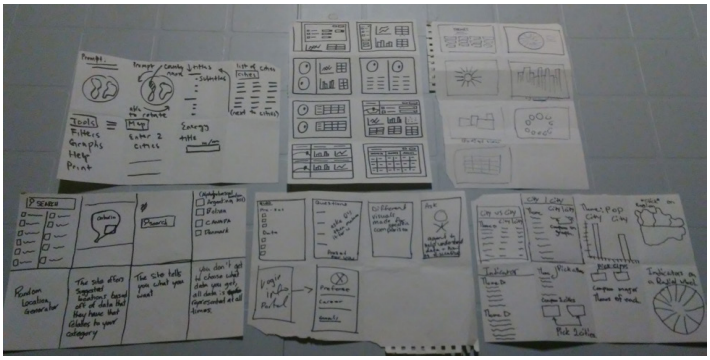
Infographic
- Shown to be the most well received (3X more shared than any other)
- Appealing and easy to understand
- Uses pictographics minimal text

Pie Chart
- A lot of people don't like pie charts but if used properly they can do a very good job at displaying the data
- Need to limit the number of sections
- If one section is 24% and one is 26% there should be clear differentiation

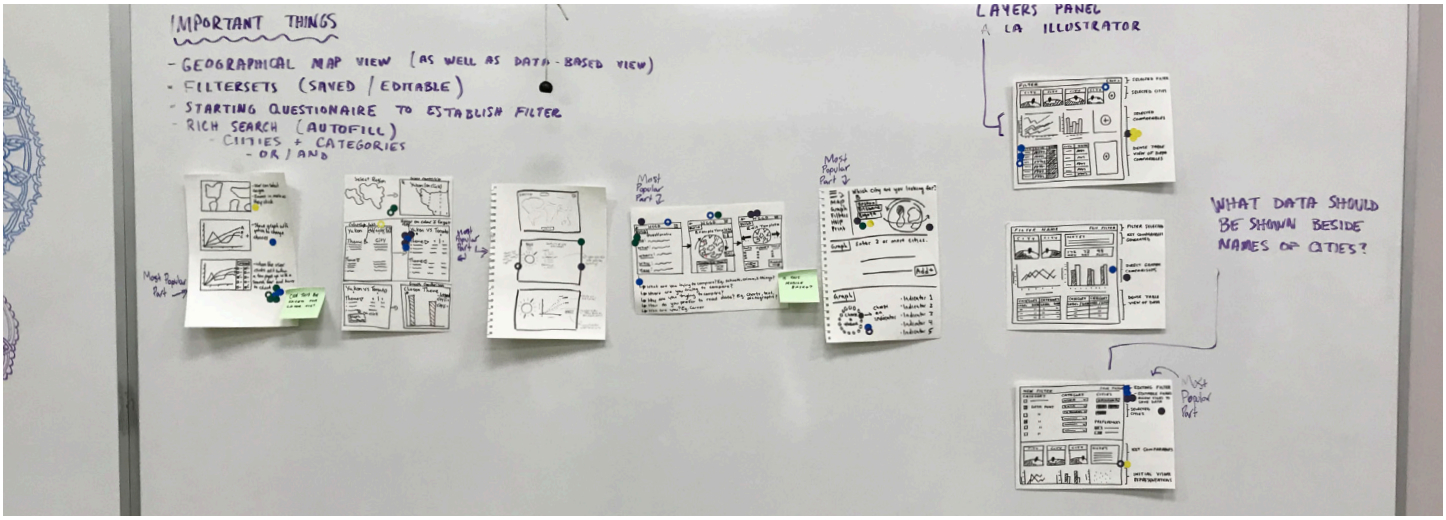
Timelines
- Meant to show progress through time
- Not fancy but it's very easy to understand



1.1 - Four Step Sketches



1.2 - Crazy 8s

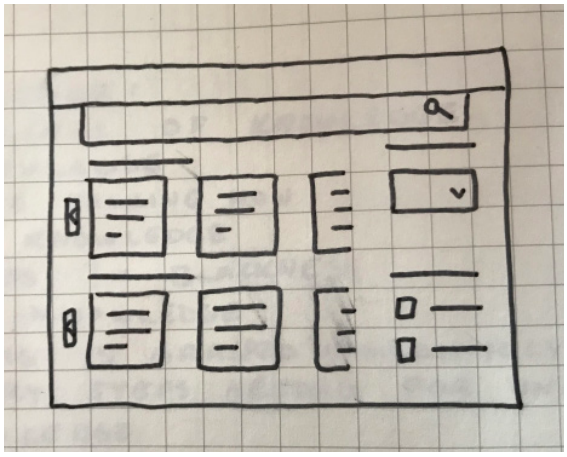


1.3 - Solution Sketches

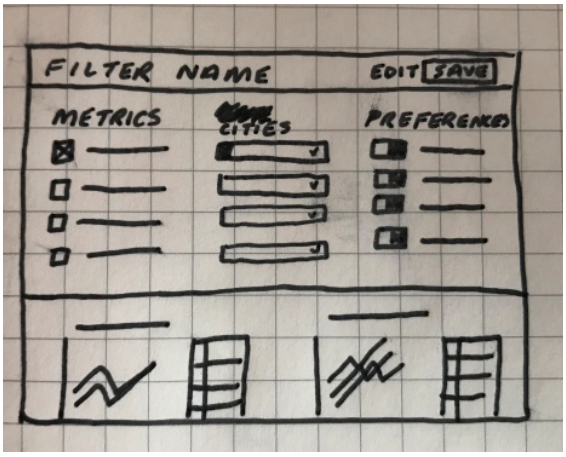
For ideation, our team conducted a four step sketch consisting of gathering notes, gathering ideas [Figure 1.1], tweaking those ideas with the "Crazy 8s" [Figure 1.2], then doing a solution sketch [Figure 1.3]. Each member reviewed our previously done competitive analysis and research material (including websites similar to WCCDs and other big data visualization sites) in our notes before moving on to generating ideas. For twenty minutes during idea generation, each member individually jot down ideas incorporating them in the form of graphs, diagrams, or anything that came to mind during the exercise they thought would be valuable to represent visually with a wireframe.

During our thirty minute solution sketch, we each drew our best ideas in the form of a storyboard consisting of 3 panels. After taping all of our finished solutions up on a board (similar to an "art museum"), we conducted a heat map where we placed dot stickers beside the parts of each solution sketch that we liked and wrote any questions or concerns on a sticky note that was then placed under the sketch.

Moving on, we went through a speed critique where we discussed each solution sketch to decide what should be included in the storyboard and prototype. We took the important parts from our solution sketches and weaved them into a storyboard [Figures 2.1-2.6], imagining how the user would find the website and then what it would look like for them when going through it. This helped us to identify gaps and priorities moving into our prototyping phase.

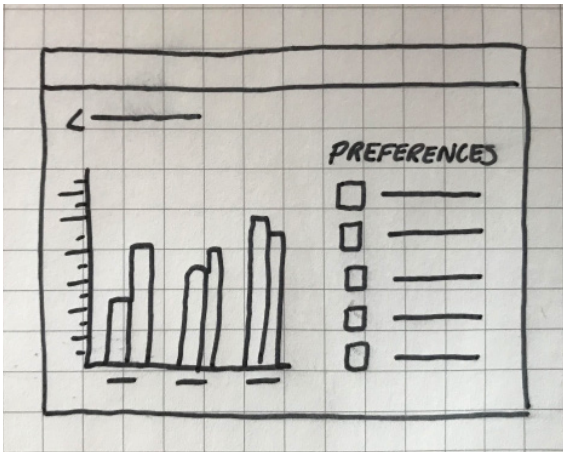


2.1 - Popular Metrics / Cities (Onboarding)

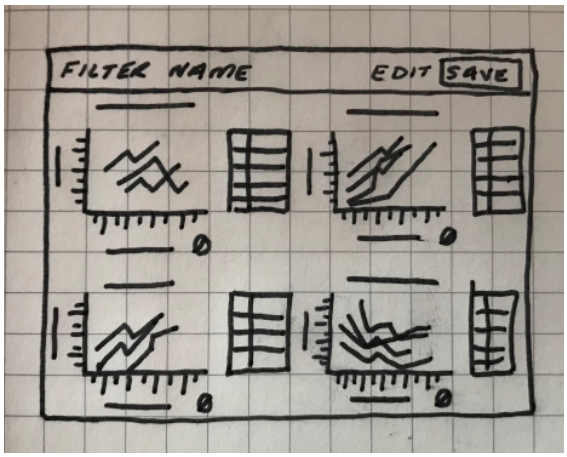


2.4 - Editing Filterset Criteria

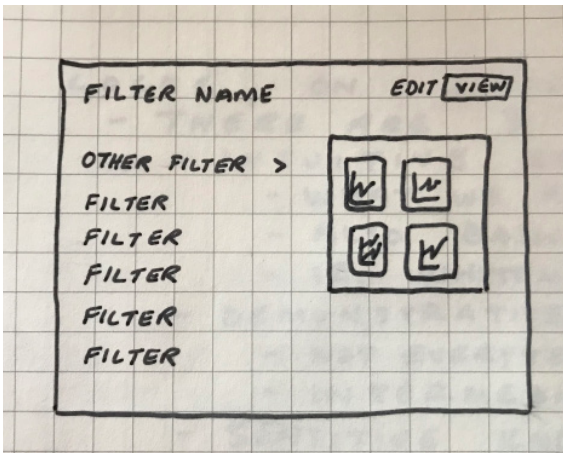
2.2 - Select Metric & Visualization Method



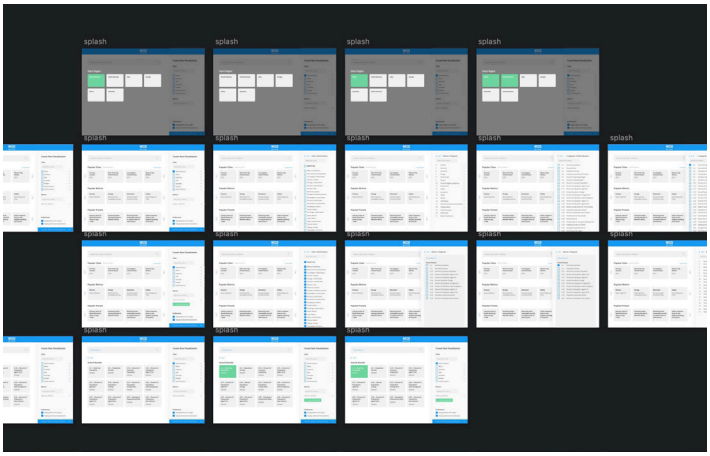
2.5 - Editing Individual Visualizations



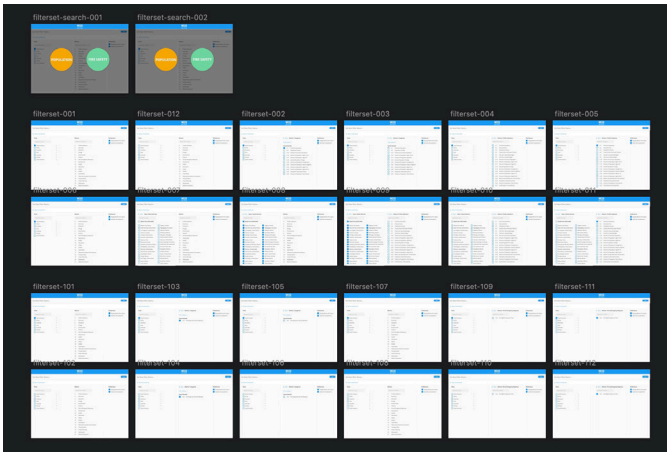
2.3 - Full Dashboard Display



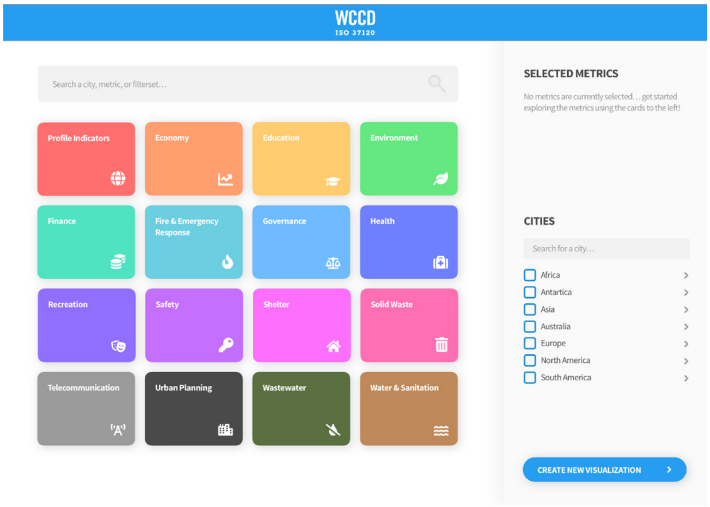
2.6 - Selecting Previously Saved Filtersets



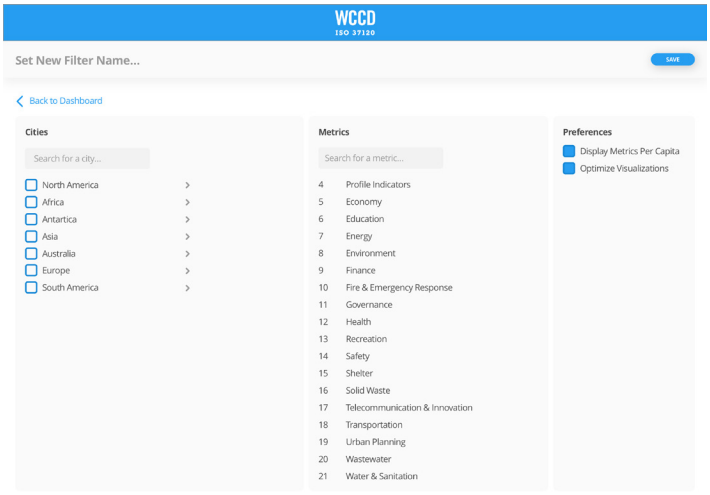
3.1 - Wide Angle of Initial Prototyping Web



3.2 - Wide Angle of Initial Prototyping Web



3.3 - Onboarding / Creation Rework



3.4 - Filter Selection Menus Rework

Some of the things we wanted to address in our first prototyping exercise was the ability to save filters for continued use at a later time, view more data at once while selecting the filters with which to view the visualization (Figure 3.4), as well as a more engaging onboarding experience (Figure 3.3). We built out a web of over 50 artboards to make sure that our users felt as though the prototype was functional... pay no attention to the Sketch file behind the curtain (Figures 3.1 and 3.2).

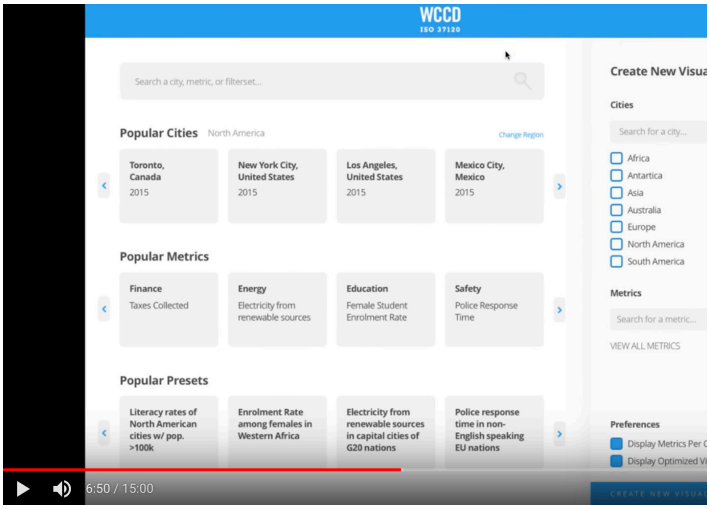
A key takeaway from our talks with contacts at Focus21, our research and ideation was the idea that we needed to incorporate richer, more accessible filter access. We drew out a number of solutions as to how that might work, eventually looking to maintain some familiarity and context of the current site by sorting the data into the subcategories laid out by ISO 37120; although our first draft of this selection process was a bit rudimentary, it served it's functional purpose to be tested. As we considered how to sort this data, density was on all of our minds, and although we could do research and make assumptions about the level of density users would feel comfortable with, there was ultimately only one way to find out, and so off we went to test our prototype!

After the completion of low to high-fidelity prototype, we were able to begin user testing. Our methods of usability testing were interviews, contextual inquiries and A/B testing. We completed our user testing with five users- three Laurier Digital Media and Journalism students and two Laurier Faculty members. Due to the constraint of access to end users such as government and city planners, we found users that were closely linked to our end users in that they regularly use data visualization websites for school and work. We conducted contextual inquiry and A/B testing by having the user complete predetermined tasks on each site. These tasks were timed, recorded by a screen cast as well as their voices were recorded as they were to do think aloud as they completed each task on each site. This allowed for us to observe how many clicks each task had per website, the level of difficulty and the user's overall thoughts and reactions to using both platforms. The feedback we received was surprising in some ways, validating in others, but obviously beneficial to the development of the project.

Key Takeaways

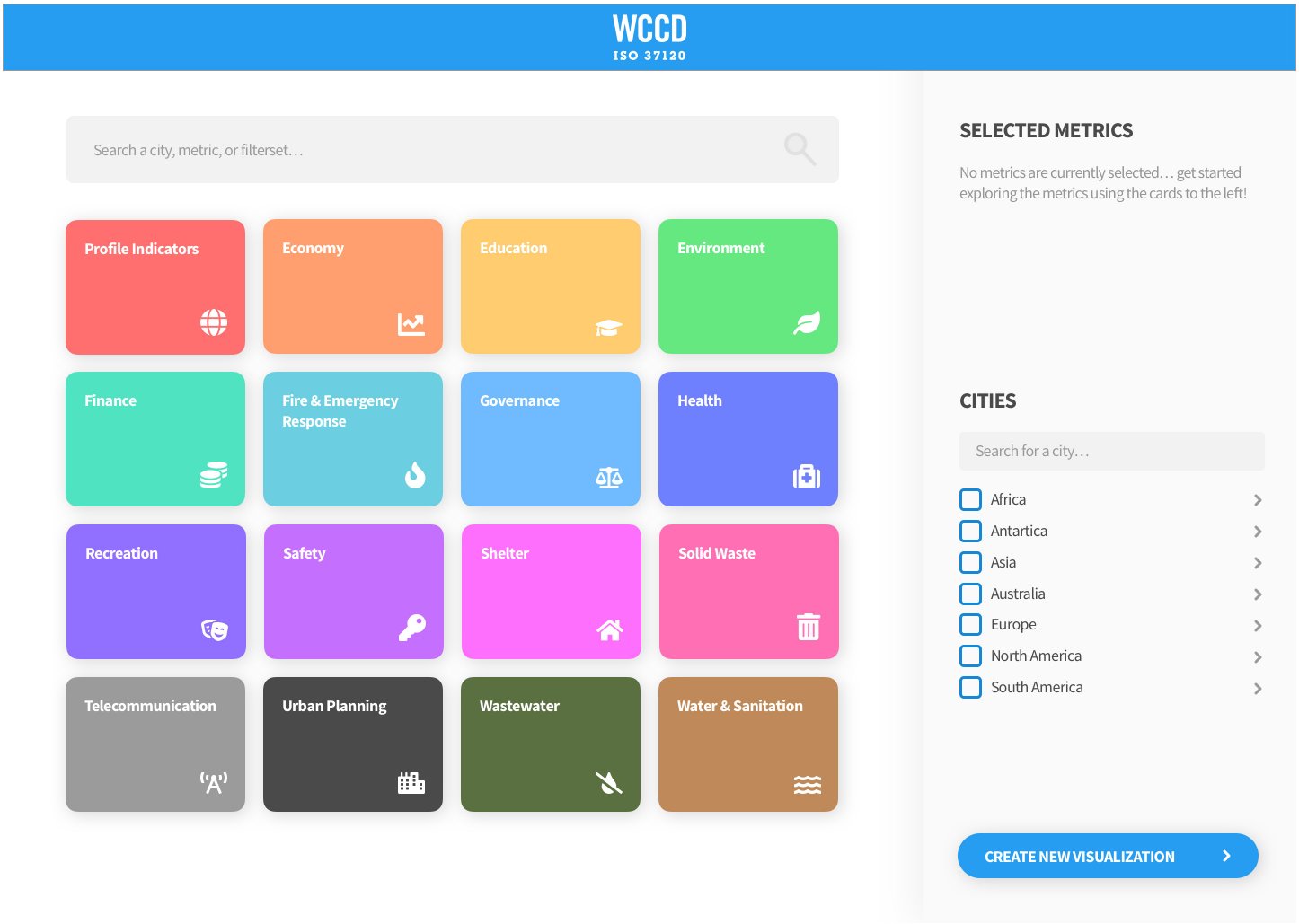
- The "number" labeling system (aka the ISO section headers) weren't effective ways of displaying metrics
- Users valued bright and captivating visuals, and would prefer spending more time interacting and learning a platform over having dull visuals*
- Navigation needed to be as simple as possible, and less clicks per task are preferable
- No users utilized the search feature in our testing

* Our initial plan for our solution involved adjusting the scatter-plot layout of the data as presented in the current site visualization. We heard overwhelmingly from the users that the way the current data was visualized was actually "refreshing," despite it's usability issues, and we elected in our final prototype to focus more on adjusting the usability issues we found through testing as opposed to the actual data visualizations. We recommend looking into the data visualizations in a further sprint -- see Recommendations on Page 18 for more information.



3.5 - Screen Recording of Onboarding Rework Testing

3.6 - Screen Recording of Dashboard Rework Testing



4.1 - Redesigned Onboarding View

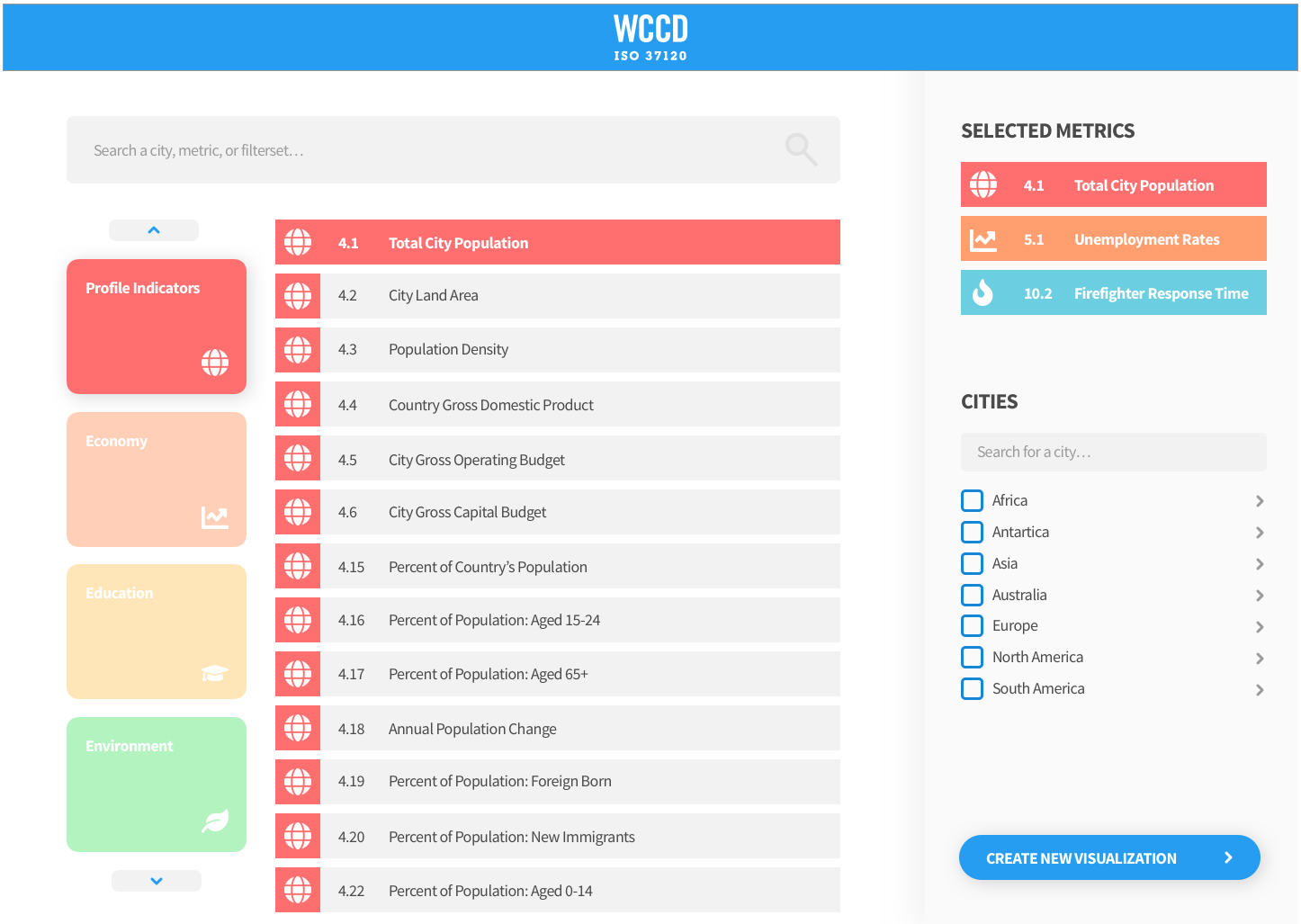
We adjusted our onboarding experience accordingly, trimming the fat off of our initial prototype and taking into account the needs and wants of our users.

Some things we brought over from the current site that worked well for our users in testing:

- Users have colours and icons work as identifiers for each of the ISO standard categories.
- Users are able to see their selected metrics at (essentially) all times during their use of the site.
- It takes two clicks (and some light scrolling) to add a metric from the main screen, down from about 4 in our first round of usability tests for our prototype.

Some key changes made from the current site to now:

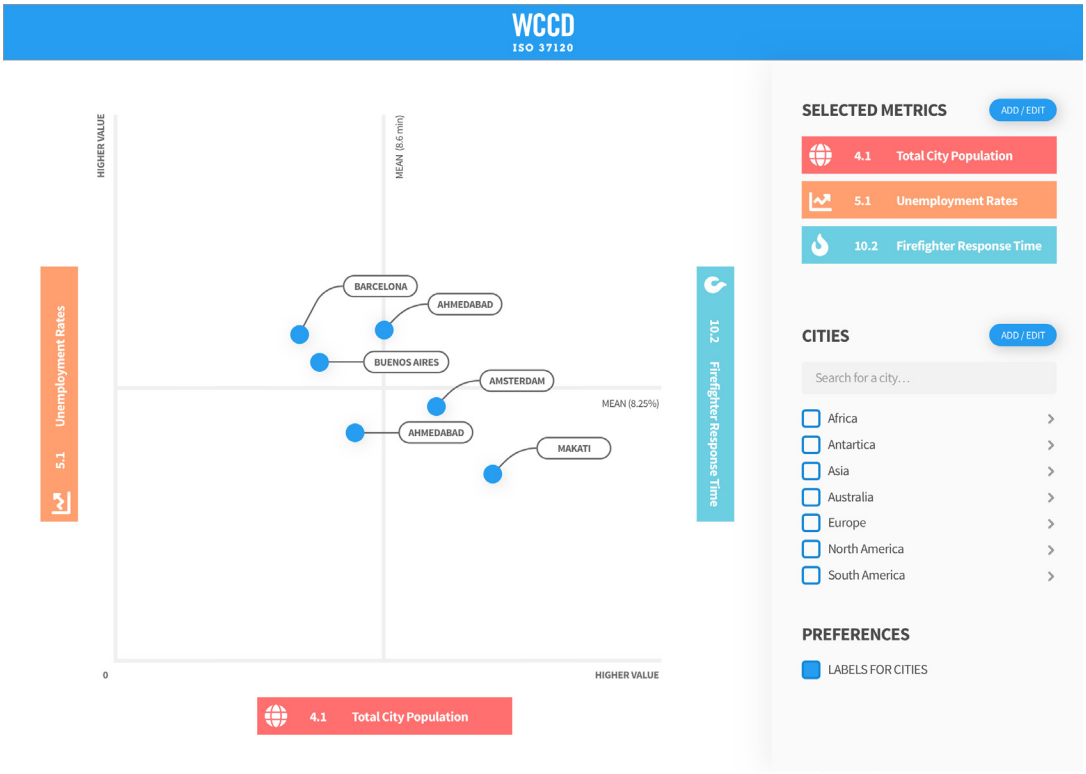
- Users are presented a list of categories with all category and metric names available to them at once, as opposed to having to scroll over or click the icons to view the category and metric labels.
- Users aren't forced into using an arbitrary, circular layout of these categories.
- Adding cities and metrics can be done via search functionality, as well as a clearly organized list of cities with multiple sorting options.
- Tasteful drop-shadow effects on white (and slightly off white) backgrounds help users differentiate portions of the site, as well as indicating intractable elements.



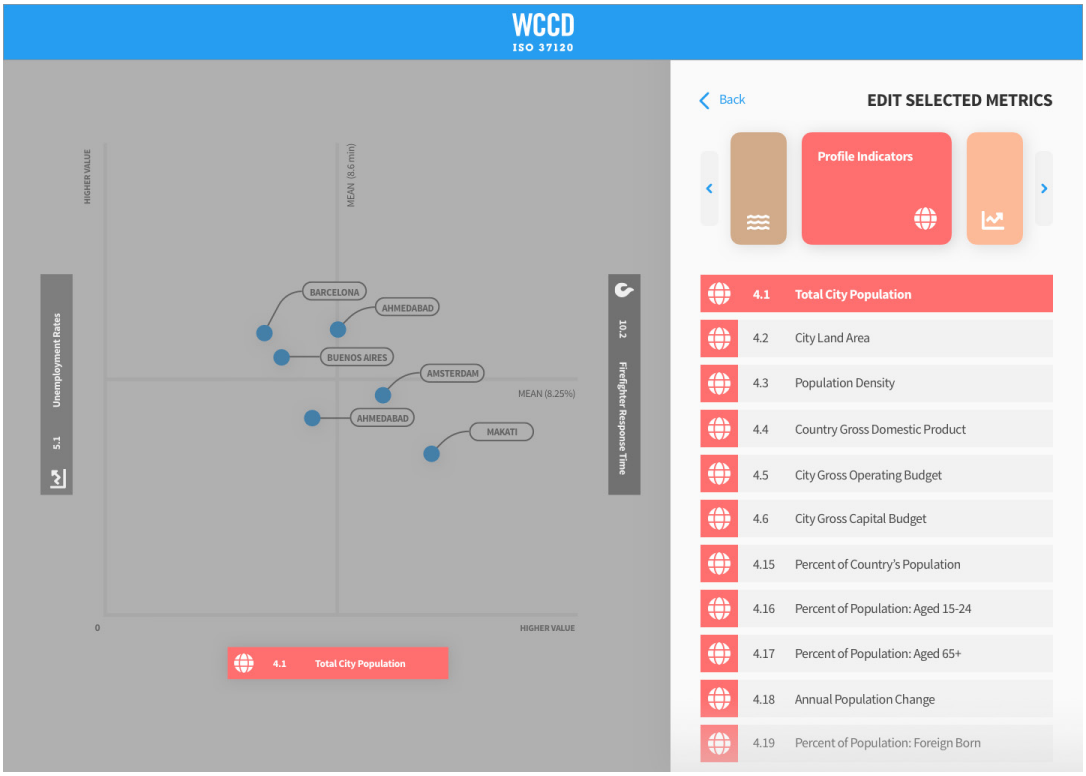
4.2 - Redesigned Metric Selection View

Some key points regarding the redesign of the site as a whole (Figures 4.1-4.4):

- Users are presented with the ISO Standard metric labeling scheme on each individual metric, but in a much less dependent way than previous prototypes.
- Users are presented a small icon and colour to maintain visual consistency across all instances of metrics and category cards on the site.
- Users are able to scroll through the list of categories and list of metrics in individual columns (in opposition to full-page scrolling), maintaining a balance of visual density for usability and aesthetic considerations.
- The selected metrics section live-updates in real time, displaying this key info regardless of how far down the rabbit hole a user may find themselves, and can be clicked to edit regardless of a users place on the site.
- The category cards stay the same size and shape regardless of your place on the site, reinforcing a consistent visual language.
- City labels can be turned off, and are all sorted to the right by default when viewing a visualization (to avoid the mess that the site currently has in place when it comes to labeling data on the scatter-plot).



4.3 - Redesigned Visualization View



4.4 - Redesigned Edit Selected Metrics View

Recommendations

- Continuing to explore the ways complex data gets represented is a key to the long-term sustainability of the site. The people we tested with generally understood the language of the city-based scatter-plot, but problems like labeling and density still exist in the current visualization. While we initially planned to tackle that challenge, our understanding of our space quickly shifted to being able to access information (selecting filters) once we realized that was a key roadblock that came before the visualization in the process. In the future, it would be worth exploring the ways to display information in a way that works in more dense situations.
- As mentioned earlier (during the Initial Prototyping stage) we attempted to sort out an option where a user was able to log in and save a list of filters for later reference, but simplified our prototype for the sake of testing our core functional changes to the systems usability as opposed to this change. Moving forward, it may be worth re-visiting the ability to save and select visualizations from a user profile or list.

In Closing

Sprout would like to thank Focus21 and the World Council for City Data for being gracious partners, and for allowing us the opportunity to work on a project that challenged us as students and designers in new and exciting ways.

We would also like to thank Seul Lee, Filip Jadczak, and Abby Goodrum for their continued guidance and support throughout the duration of the project, as well as our testers from the Digital Media and Journalism program at Laurier Brantford for being kind enough to let us test our ideas on them.

We wish Focus21 and the WCCD all the best moving forward with the project!