Process Book

Claire Pang & Vi Le

Overview & Motivation

When you are asked to think of skyscrapers, your immediate thought may be of a famous building. From the Empire State Building in NYC to the Space Needle in Seattle, they are all very grand. But, can they tell us anything about the communities they are located in?

With the amount of free land space shrinking in the United States, buildings have been building upwards rather than outwards. This is an increasing trend, if not problem, in metropolitan cities such as San Francisco and New York. Likely, skyscraper construction will continue in order to meet housing demands, which will play a role in reshaping the iconic skylines of major cities.

Skyscrapers are generally filled with offices. With an increase in the number of workers in the area, we also expect an increase in the number of residential buildings. Residents, in turn, need something to do besides go to work. This can be anything from movie theatres to museums to casinos. These major centers also attract non-locals into the city who then require hotels to stay in. We would like to examine the evolution of skyscraper construction in the United States through an interactive data visualization.

Project Objectives/Questions:

When we first started our project, we were going to focus primarily on US skyscrapers (because we found a dataset that had that readily available). Therefore, we focused on these questions:

- Which are the tallest buildings in the US? (and how tall are they, relative to other tall buildings in this dataset?)
- Where are the tallest buildings in the US? Does a certain city/state have the most tall buildings?
- What are the purposes of these skyscrapers (casinos, offices, hotels, residential, etc)?
- Is there a constant positive trend for the amount of skyscrapers constructed, or has construction halted?

However, once we finished answering those questions, we decided to expand and briefly examine other skyscrapers in the world, leading to the following question:

- How do US skyscrapers compare to other skyscrapers in the world?

Data:

Initial processing of the data from CORGIS:

- Adding a “state” column (sorting by state, then manually entering/researching)
- Removing duplicates (if a building was rebuilt)
- Sorting the file by highest to lowest height (Excel)
- Adding some more into “observation”
- Adding missing heights and/or locations
- Removing buildings with height < 130m

The original CORGIS dataset needed to be cleaned up since they were missing some values that we needed for our visualizations. One of the problems that we encountered was that I encountered was that there was no “state” column. Since different states have cities with the same name - for example: Albany, CA vs Albany NY, I had to first alphabetize by city, then use google to search by city and cross-referenced that with building name. Sometimes the building building name did not exist. For this, I used google to search by the building name with no city. Since I had to go through each unique

individual city, I simply created a new column using google sheets since it didn't make sense to code a new state for every new city.

Another problem with the original dataset is that it had some missing coordinates. First, I searched for building name. When the building name did not give me an address and coordinates, I used skyscrapercenter.net to search for the building name. This gave me alternative names that I was able to use to get an address and use google maps to enter an address to get the coordinates.

After going through the dataset, we realized that some of the purposes of the skyscrapers were missing. For example, the Stratosphere Tower did not have "Observation" as true. Since this affected the validity of the dataset, we went ahead and used google to verify their purposes. The problem was exclusively in Las Vegas. If there was a city that is known for gambling, we made sure the proper purposes was listed.

After working on some of the visualizations, we realized that considering all of the skyscraper's purposes (example: if a skyscraper is both residential and hotel, or residential and office) may be misleading and potentially double count skyscrapers. Therefore, for simplicity, we decided to give each skyscraper a MAIN purpose. This required using data from skyscrapercenter.com and then processing it (CSV we used can be found in our github repo under data/skyscrapers-uses.csv). If a building is multipurpose, the purposes are listed as "main purpose/other purpose…" therefore, we cleaned up the data to include only the main purpose.
This processed dataset is in our github repo under data/skyscrapers-main.csv.

For the line chart, the data was processed in R so that the line chart could be made more efficiently. Basically, the number of skyscrapers completed in a year was calculated, and then added to a cumulative count. This was additionally done for each of the main purposes: hotel, residential, office, and other.
The script that calculates this can be found in our github repo under code/yearProcessingCount.R.
The dataset that we used for the line chart (result of the yearProcessingCount.R script) is available in the github repo under data/skyscrapers-count.csv.

For the bar chart, the data was processed in R as well to find the Top 10 Tallest in every city that has > 10 skyscrapers by 2016. The code that does that is also included in the yearProcessingCount.R script (starting at line 147). The result of this script / dataset we used is available in the github repo under data/skyscrapers-top10.csv.

For the world skyscrapers visualization, the data (name, location, height) is from Skyscraper Center (where the data is originally scraped from before it went onto CORGIS).
Specifically, the link to the page is here.
The data was manually entered/scraped for the top 20 (with some alterations, because we did not want to include the Petronas Twin Towers as 2, or the buildings that are not completed). The dataset is available in our github repo under data/world.csv.

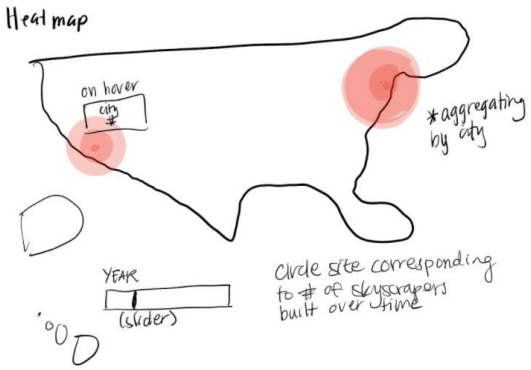
The images (renderings/drawings of each of the skyscrapers) used in the visualization are from clicking on each of the skyscrapers on the Skyscraper Center website (see example above - the link is also here). They were manually screenshotted and then cleaned up in photoshop so that:
a) there is no whitespace
b) the picture is not too wide
which would allow the skyscrapers to be properly scaled in the visualization.
The images are in our github repo under code/images.

Exploratory Data Analysis

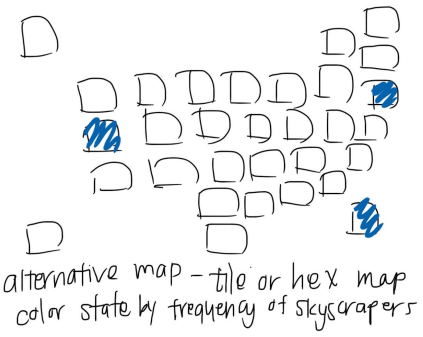
We looked at the data in Tableau and Excel, and looked at the different purposes but did not do anything extensive. If we had, this would probably have helped with creating our designs. However, as we were constantly changing the data/processing it, we did our data exploration while coding the visualizations which was probably not the best thing to do.

Initial Designs (With Related Work / Inspirations)

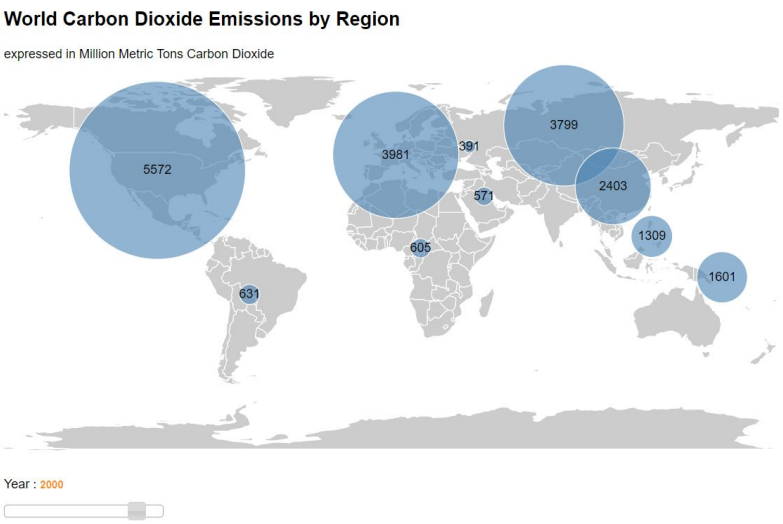
The initial designs (as done in the Project Proposal) are:



Symbol map

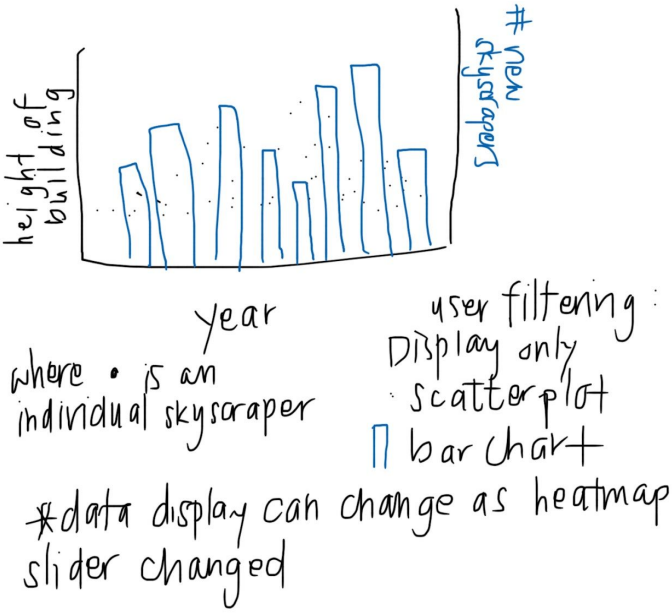


Alternative design

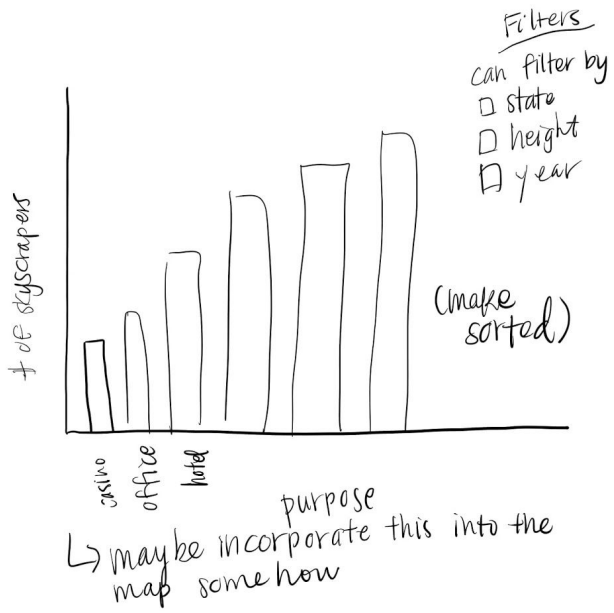


The inspiration for the symbol map was from a block.

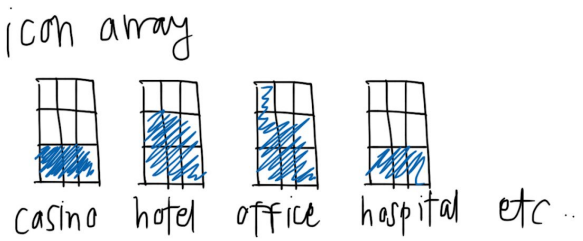
Initial design for our bar chart of top 10 tallest.



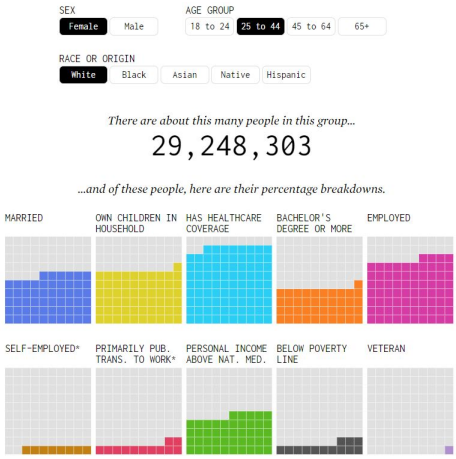
Initial design comparing # of skyscrapers built per year, along with the heights.



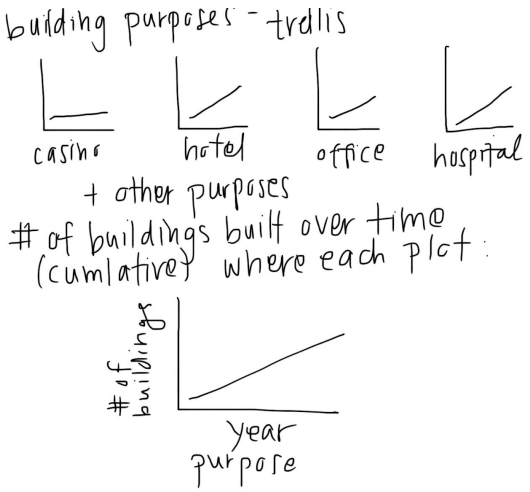
Initial design to display the purposes.



Revised design (in revised project proposal) to display the purposes. We wanted to stray away from bar charts.



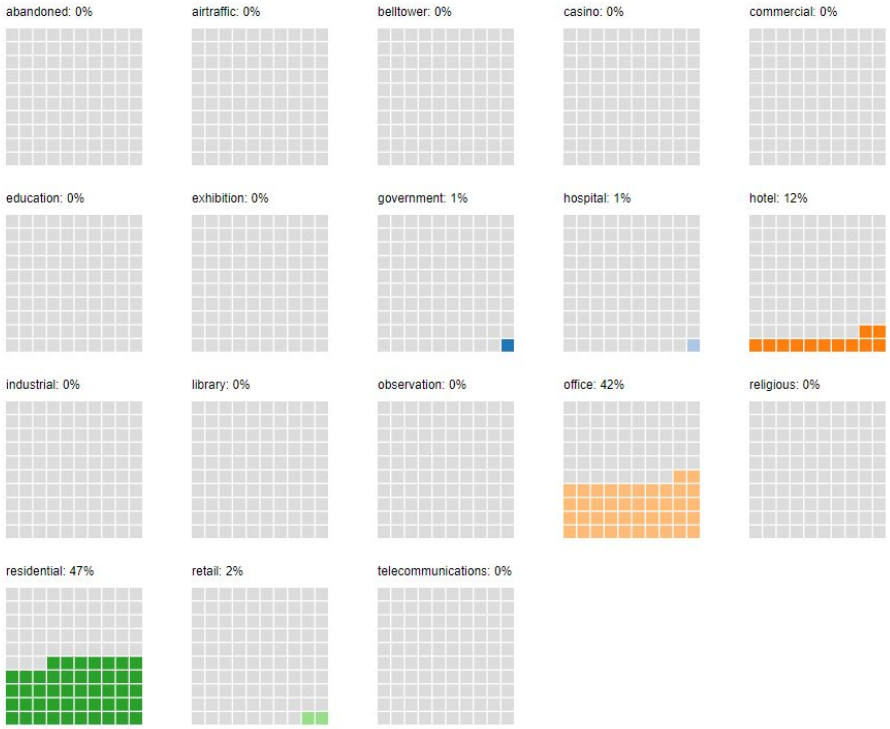
The inspiration for this was from an example by Nathan Yau of Flowing Data.



Initial design for a plot to see the number of skyscrapers and the corresponding purposes over time (we thought maybe as a trellis plot).

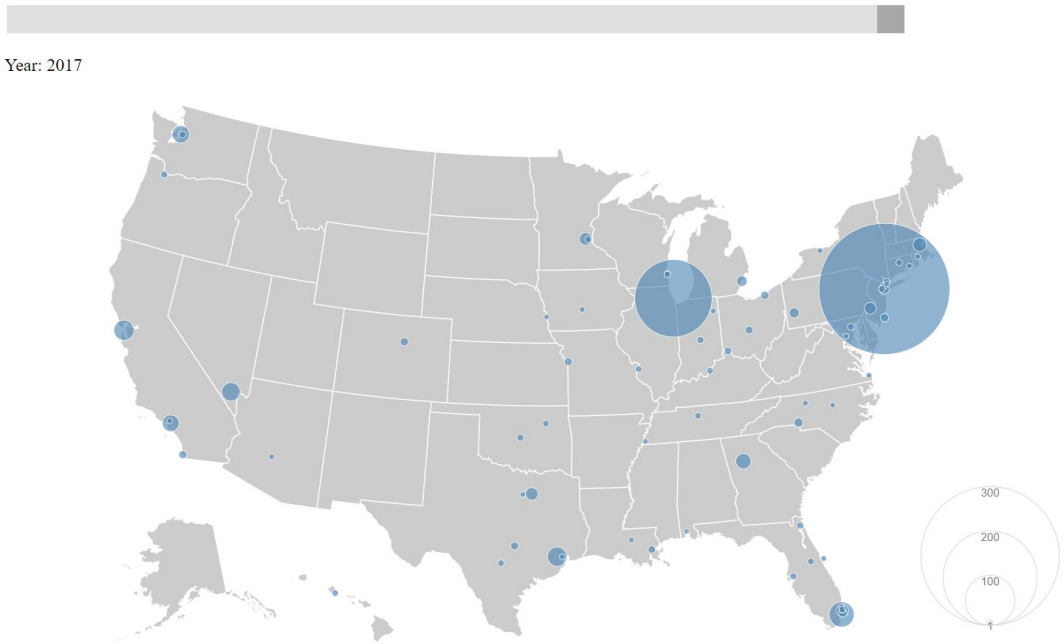
Design Evolution

APRIL 14



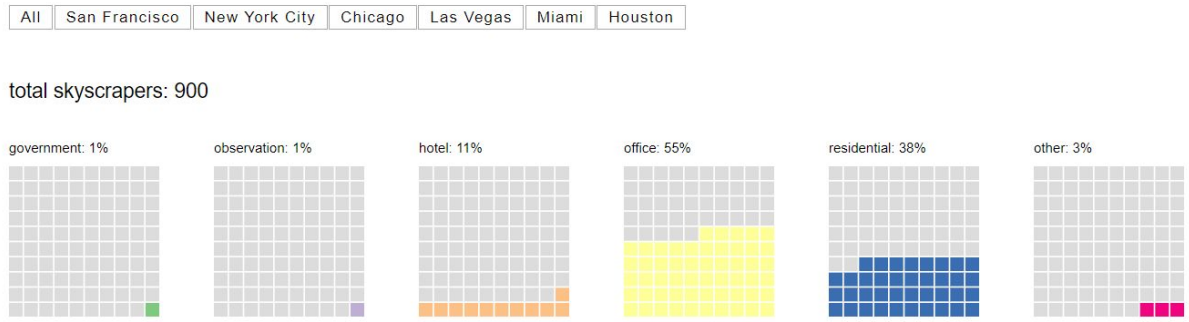
Note that this is BEFORE the data has processed with the 'MAIN' purpose column.
The first attempt at the icon array. All purposes are given their own icon array, but obviously there are not enough out of total skyscrapers to even total 1%.

APRIL 16



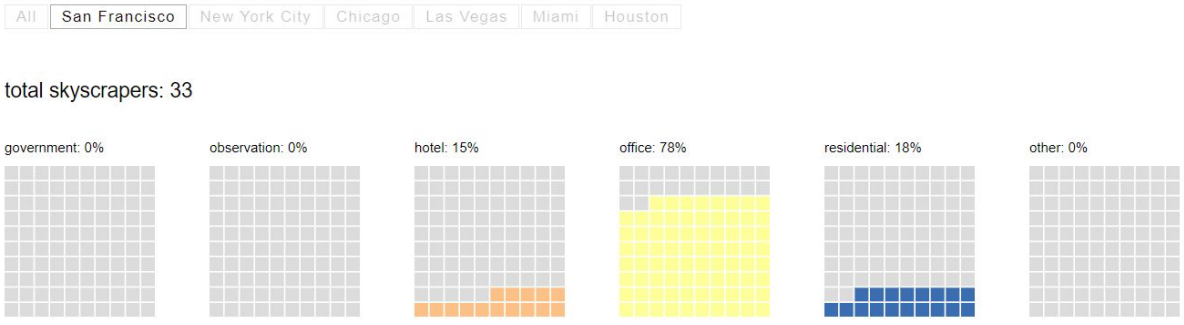
First attempt at the symbol map was a success.
At this point, div.html tooltips are being used.

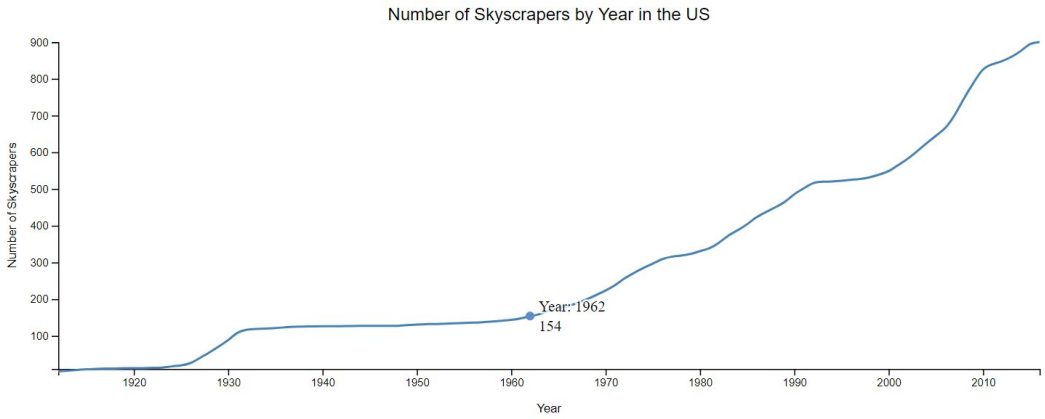
APRIL 17



Revised version of the icon array
If a category is at least 1% of the total, it gets its own icon array. The 6 cities with the highest skyscraper count is included and able to be filtered.
Because the percentages do not add up to 100% (as a skyscraper can have multiple purposes), we are weary that this might be misleading.

Filtering by a city leads to some categories having 0%:

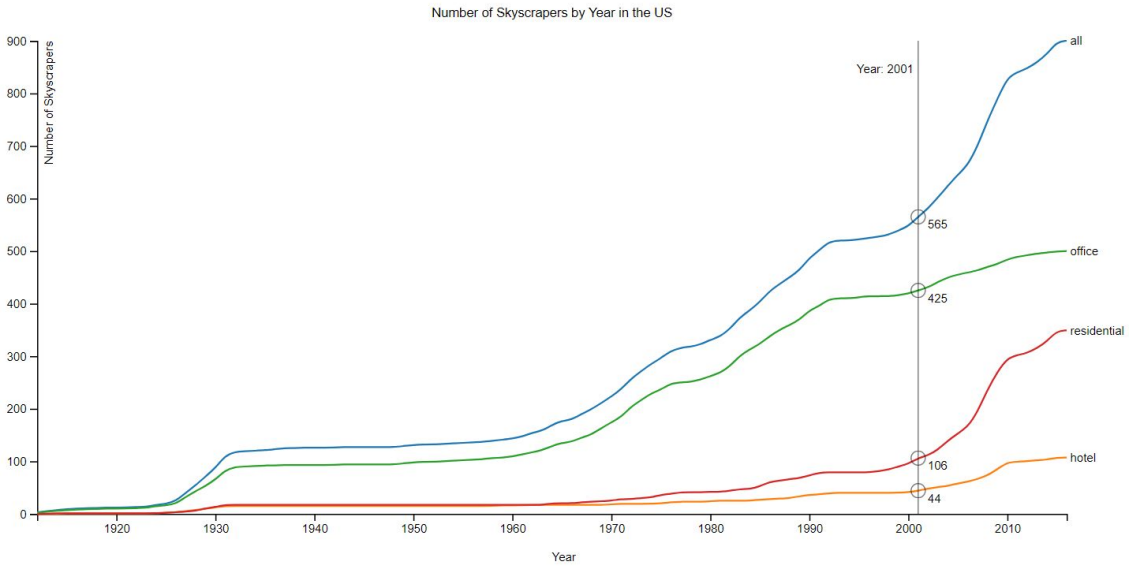




Instead of the trellis plot, we decided to do a line chart that shows the amount of skyscrapers existing in the US over time.
This is when the yearProcessingCount.R script is used to generate the cumulative count.
The tooltips were also very difficult to figure out:

- At first I tried drawing clear rectangles and using those as the area to detect where in the line it is being hovered over. It was super glitchy and flickered.
- Then I tried changing the rectangles to circles. Still super glitchy.
- I looked up other examples on how people did hover on line charts and found there is a complicated algorithm involving bisecting and finding the nearest point. Implemented it and now the interaction is smooth.
- Added text twice, once being the actual text and the other being a stroke so that the tooltip is not covered by the line.

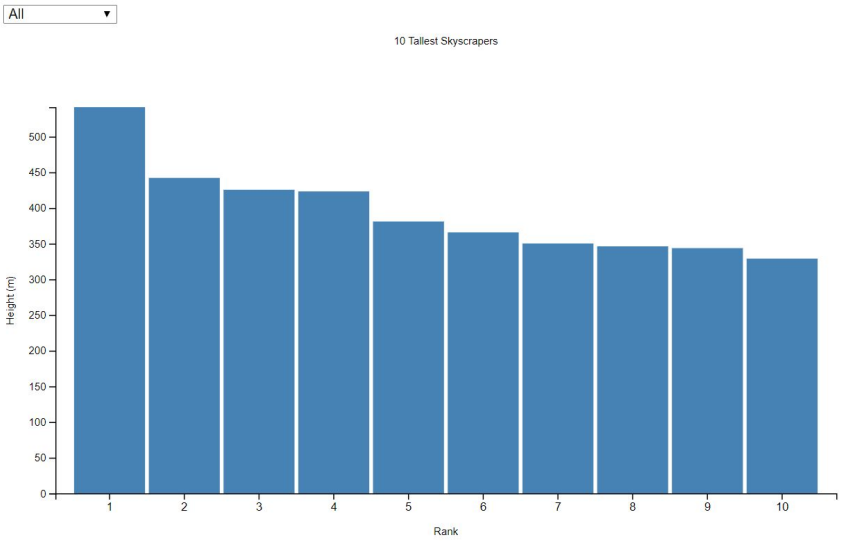
APRIL 18



Updated line chart, with categories (purposes) of All, Office, Residential, and Hotel.

- Updated data processing script in R to calculate the totals for each purpose
- Added tooltip line, filtering, and brushing.
- The “Year” line with all the corresponding values is based off this block but is implemented differently to include the actual values rather than calculating using the y-axis scale.

APRIL 21



Bar chart, with transitions and drop down menu, is done.

APRIL 23

Because we were worried about the double counting of skyscrapers to show the purposes may be misleading, we decided to add a column “main” purpose, to clear up confusion and make the visualizations simpler. (Lots of data processing required for this task).

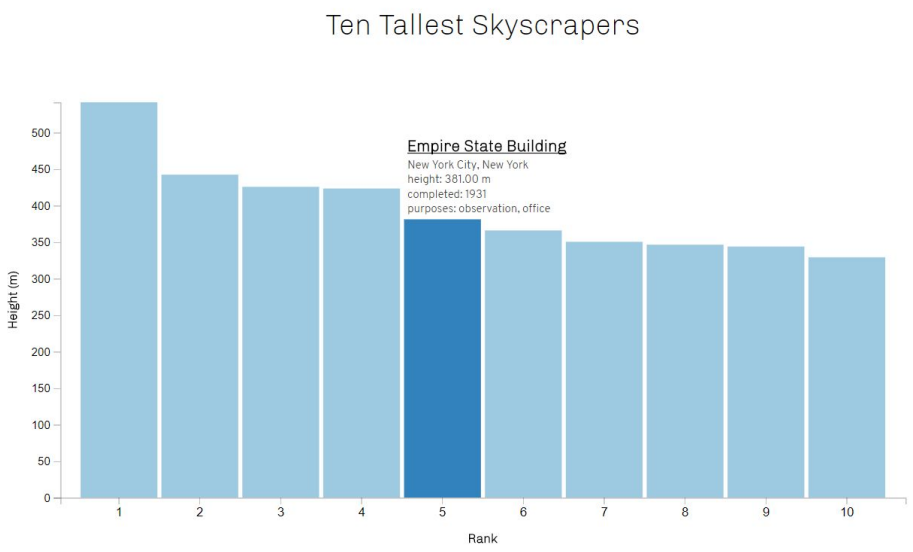
#		Building Name	City	Height (m)	Height (ft)	Floors	Completed	Material	Use
1		Burj Khalifa	Dubai (AE)	828	2,717	163	2010	steel/concrete	office / residential / hotel
2		Shanghai Tower	Shanghai (CN)	632	2,073	128	2015	composite	hotel / office
3		Makkah Royal Clock Tower	Mecca (SA)	601	1,972	120	2012	steel/concrete	other / hotel
4		Ping An Finance Center	Shenzhen (CN)	599.1	1,965	115	2017	composite	office
5		Lotte World Tower	Seoul (KR)	554.5	1,819	123	2017	composite	hotel / residential / office / retail
6		One World Trade Center	New York City (US)	541.3	1,776	94	2014	composite	office
7		Guangzhou CTF Finance Centre	Guangzhou (CN)	530	1,739	111	2016	composite	hotel / residential / office
8		TAIPEI 101	Taipei (TW)	508	1,667	101	2004	composite	office
9		Shanghai World Financial Center	Shanghai (CN)	492	1,614	101	2008	composite	hotel / office
10		International Commerce Centre	Hong Kong (CN)	484	1,588	108	2010	composite	hotel / office
11		Changsha IFS Tower T1	Changsha (CN)	452.1	1,483	94	2018	composite	hotel / office
12		Petronas Twin Tower 1	Kuala Lumpur (MY)	451.9	1,483	88	1998	composite	office
12		Petronas Twin Tower 2	Kuala Lumpur (MY)	451.9	1,483	88	1998	composite	office
14		Zifeng Tower	Nanjing (CN)	450	1,476	66	2010	composite	hotel / office
15		Willis Tower	Chicago (US)	442.1	1,451	108	1974	steel	office
16		KK100	Shenzhen (CN)	441.8	1,449	100	2011	composite	hotel / office
17		Guangzhou International Finance Center	Guangzhou (CN)	438.6	1,439	103	2010	composite	hotel / office
18		432 Park Avenue	New York City (US)	425.5	1,396	85	2015	concrete	residential
19		Marina 101	Dubai (AE)	425	1,394	101	2017	concrete	residential / hotel
20		Trump International Hotel & Tower	Chicago (US)	423.2	1,389	98	2009	concrete	residential / hotel
21		Jin Mao Tower	Shanghai (CN)	420.5	1,380	88	1999	composite	hotel / office

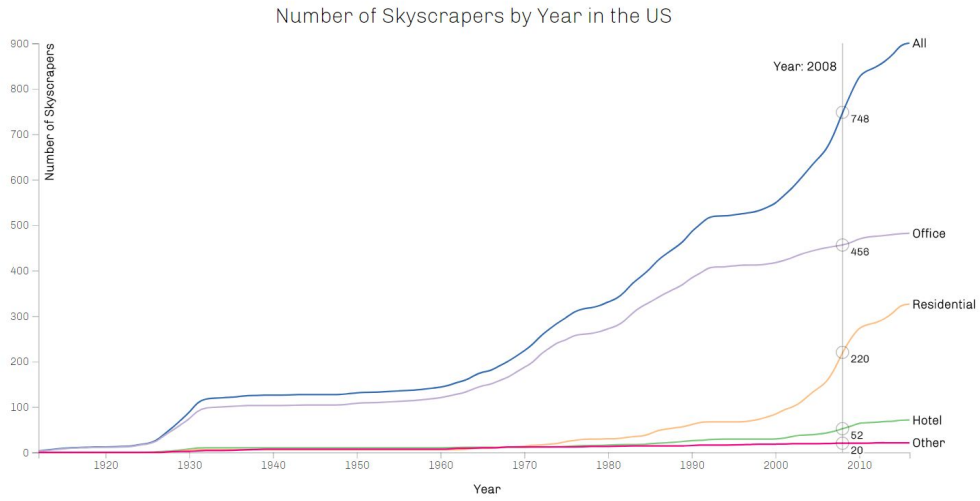
Started to look for worldwide data. This was done copy/pasting from the skyscrapercenter.com into google sheets.

APRIL 24

- Visualizations are updated to use the main purposes so that no skyscrapers are “double counted”
- Formatting (fonts, colors) of all visualizations. Colors between icon array and line chart match up

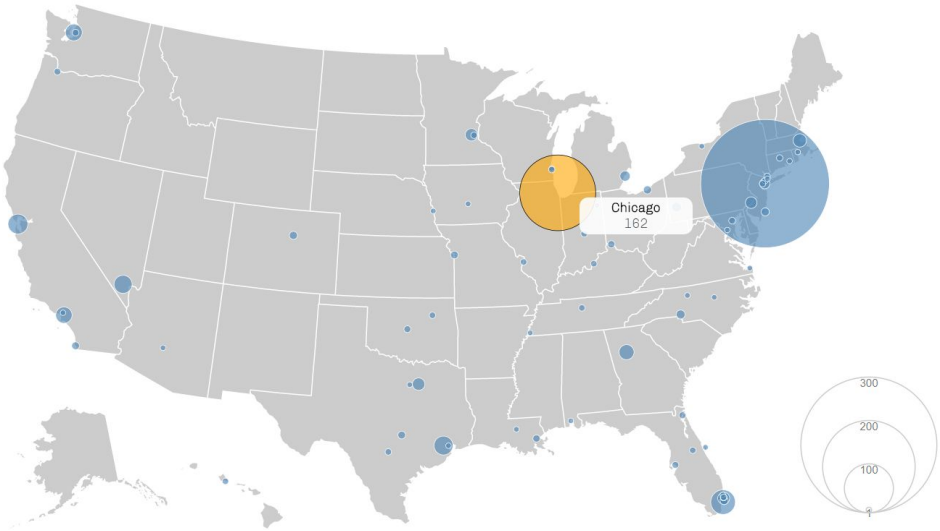
All

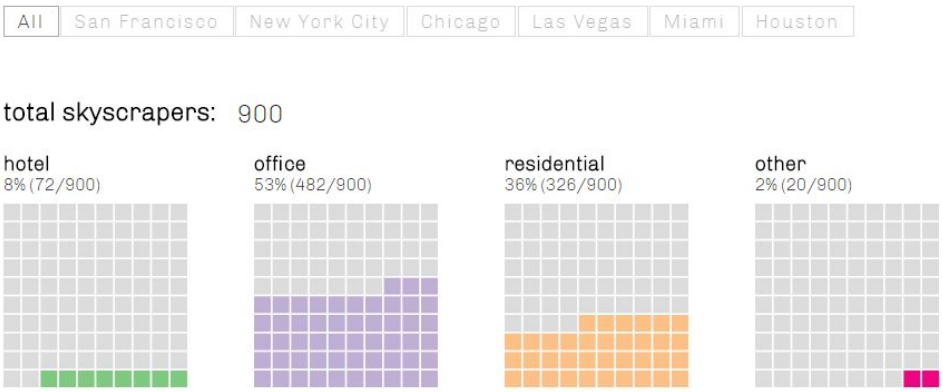




Another category of “other” is also added, so that “All” adds up to the total of office, residential, hotel, and other.

Year: 2016

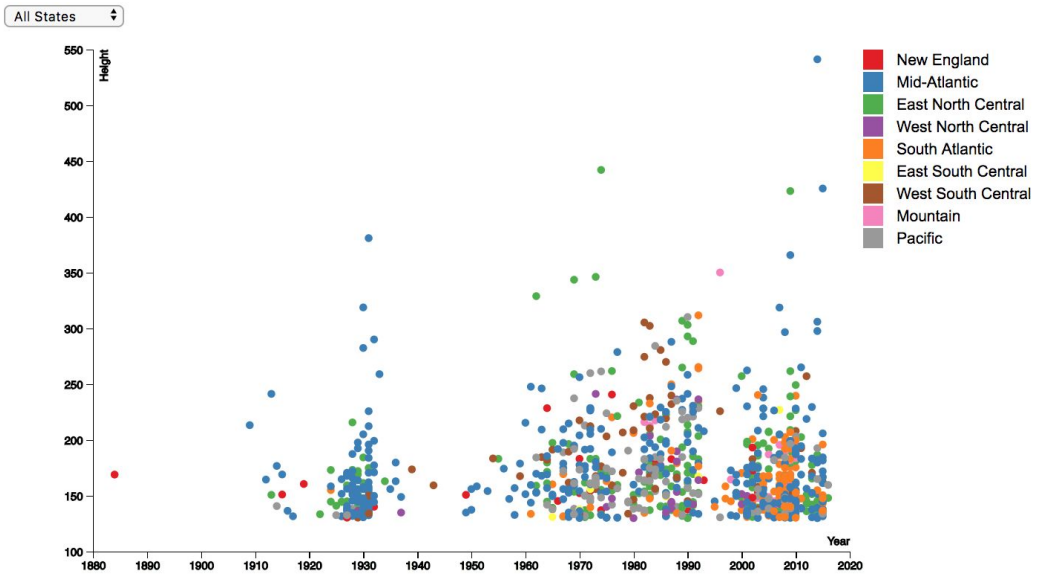




The categories of hotel, office, residential, and other are chosen as categories for the icon array.

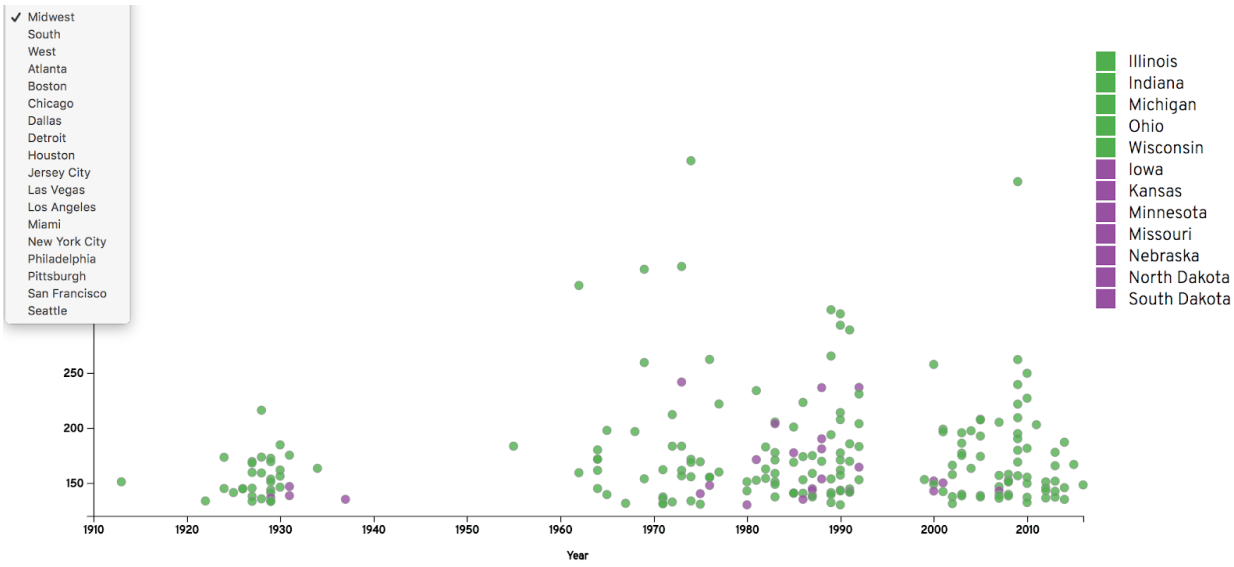
APRIL 25

- "World" visualization is created, comparing the heights of the tallest 10 skyscrapers in the world.
- The data scraping for the images + world data is also done.



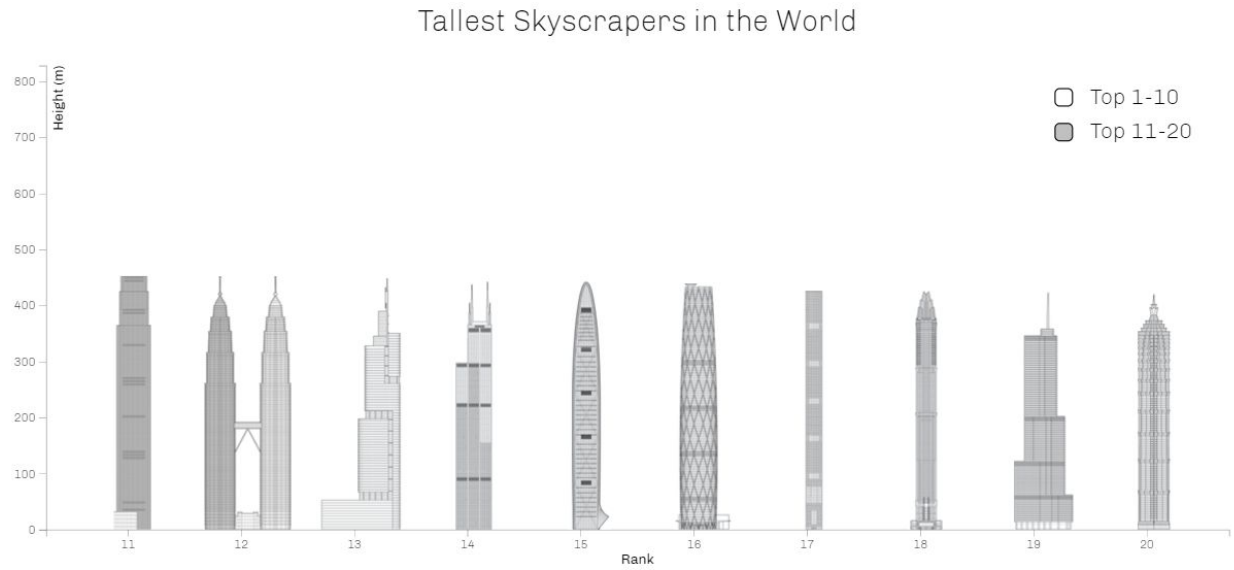
- Scatter plot created

- The multiple colors on the scatter plot did not yield much information.
- On-hover: the color of the dot did not change.

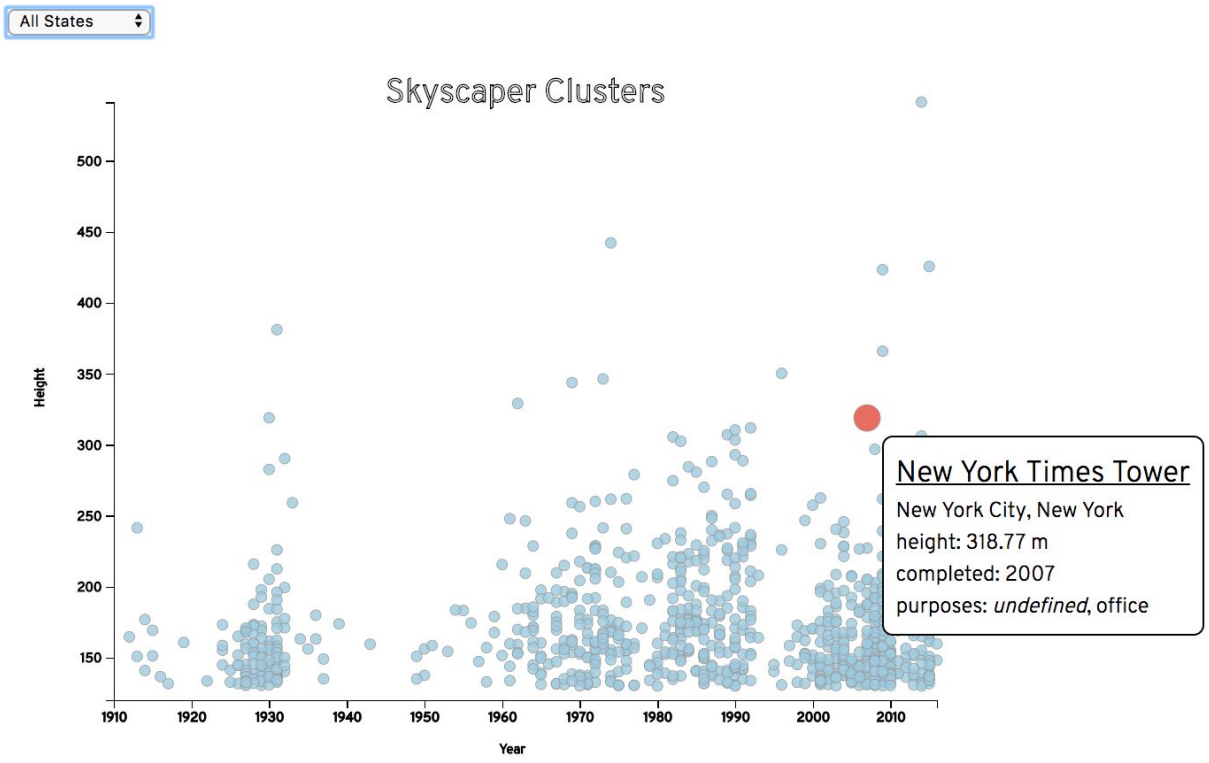


- It had the same problems as the “all-states.” There is too much visual clutter.
- Added regions. Regions were matched with the [Census Regions and Divisions of the US.]

APRIL 26



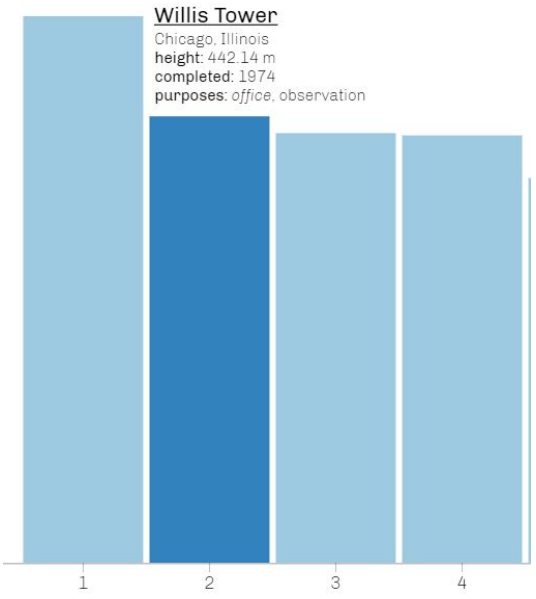
- Added ranks of 11-20 to the world visualization, with the menu on the side
- Menu was put on the actual visualization (rather than the buttons as in the icon array) because it made it more compact
- The reason why only 10 are shown at a time is because some were too wide to be scaled properly without affecting the height.

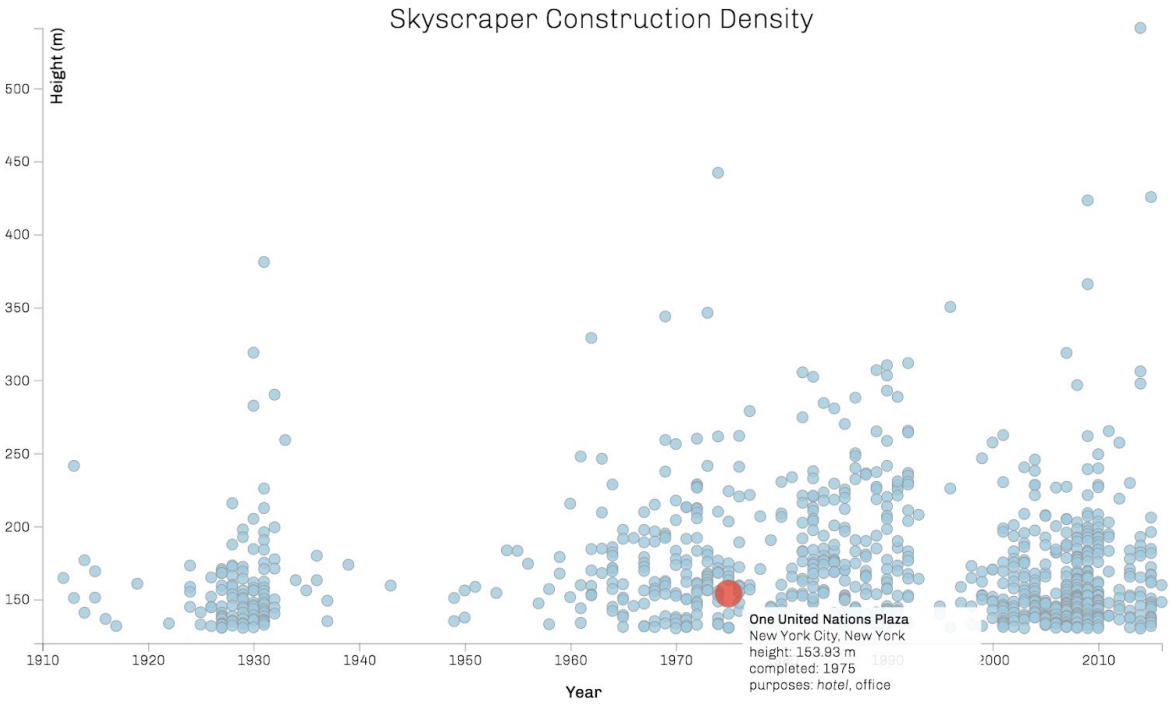


- Changed it so that the colors for the dots are the same color.
- Added title and tooltip, but need to debug why it returns undefined

APRIL 29

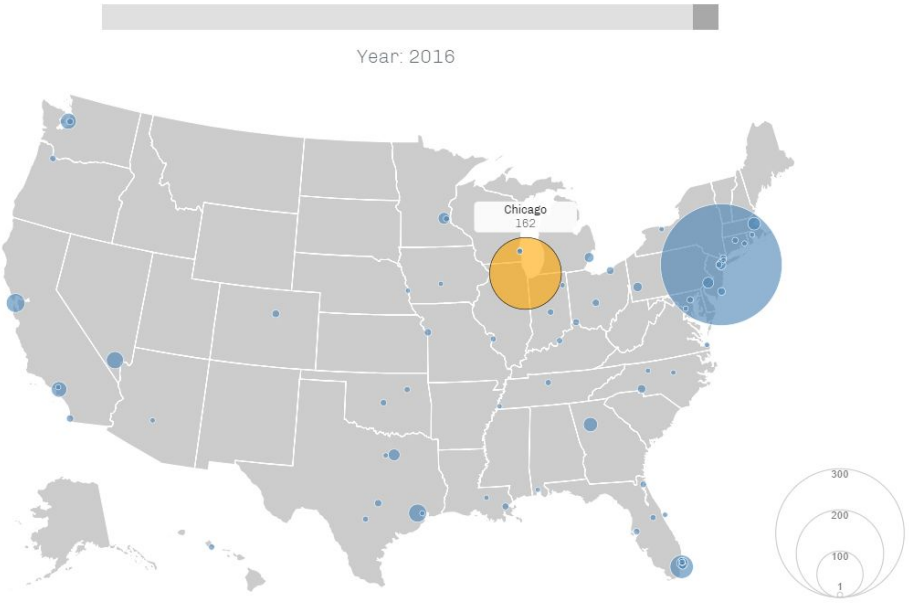
- Code edits to embed the visualizations onto github website (splitting up into separate .js file for each visualization)
- Because the div.html tooltips no longer work if it is not the only thing on the page (like in a block), we are now using “foreignObject”s for the tooltips. It allows us the flexibility of using html in the tooltips without having to deal with the problems that the div.tooltips gave us before.
- Now, we can add different styling to the tooltips for the bar chart:





- Debugged undefined. Now the tooltips purposes are information is working.
- Matched the tooltip to the other visualizations.

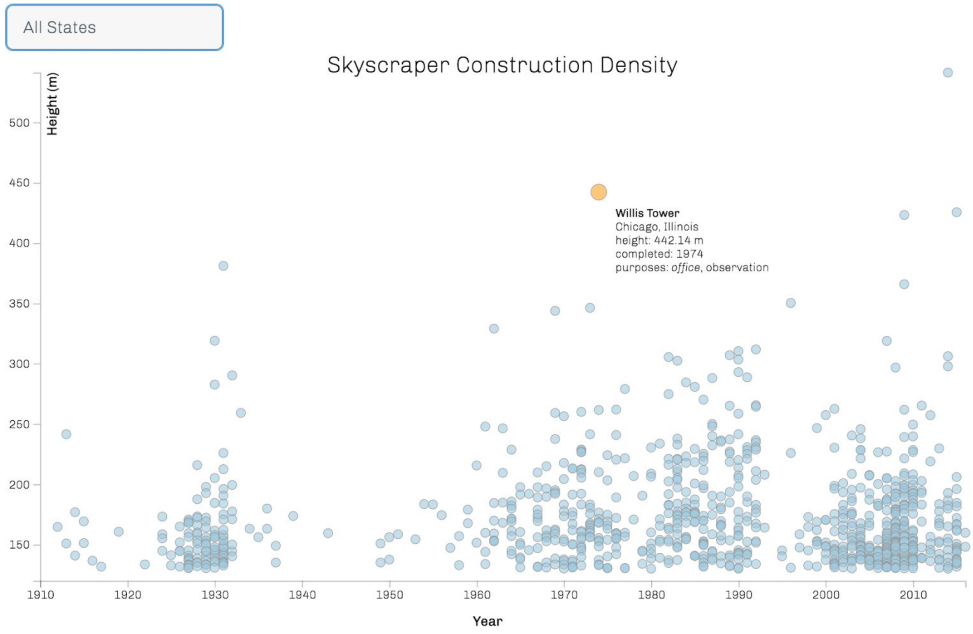
MAY 7



Edits to the slider and “Year” text

- Before, the slider and “Year” text were centered using html
- The slider was at 65%, so it would change based on screen size
- Now, the slider has a fixed width AND both slider/text have fixed position so it will always be centered with the map no matter the screen size

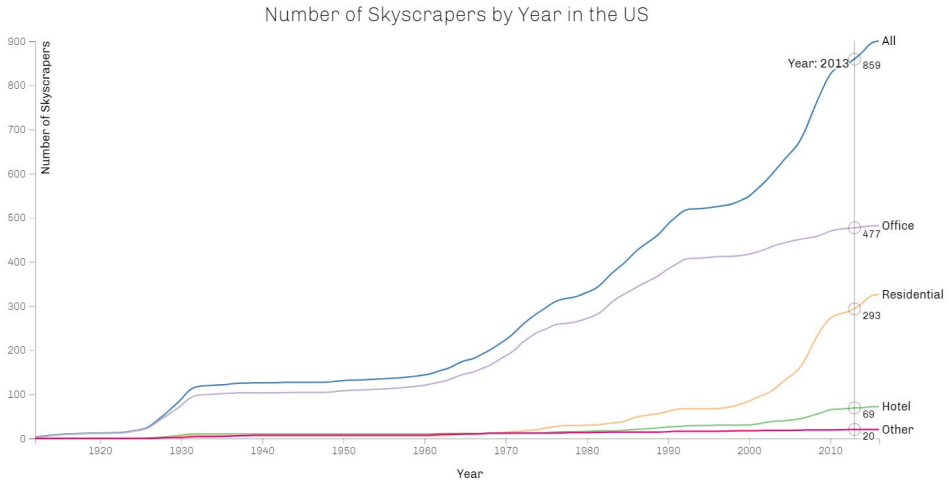
MAY 15



- Fixed the regions so that they are all the same colors.
- Opacity changed to take into account overlapping dots.
- The darker it is, the more dots there are.

Implementation / Final Designs

LINE CHART

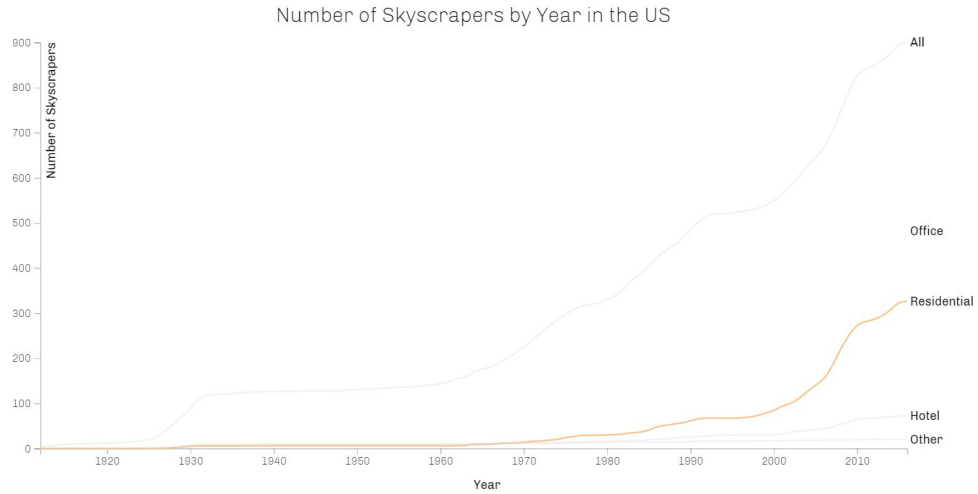


In the above example, the mouse is hovering over the chart at the x value corresponding to year 2013.

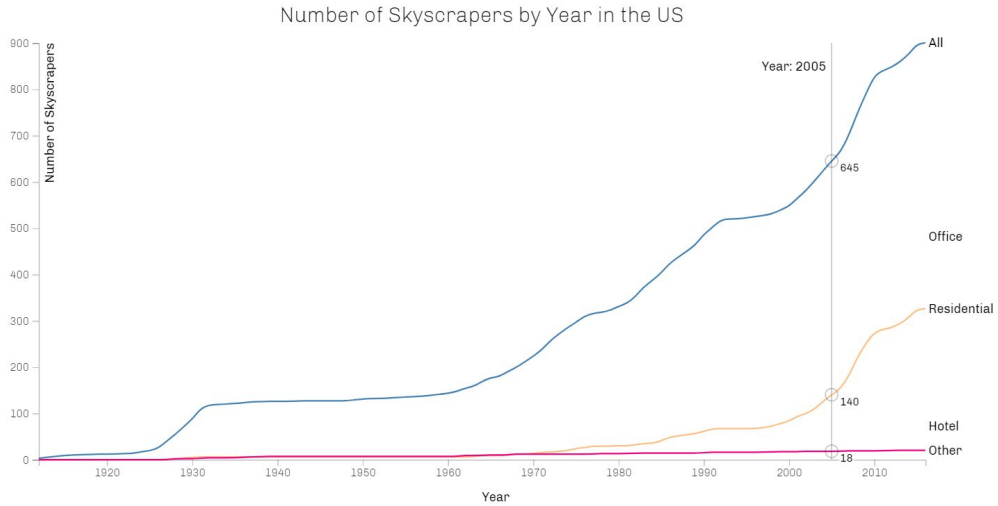
Functionality:

- Hover over the chart creates a vertical line giving more information about Year and the values for each category, for the year closest to the cursor
- Hovering over a category name (All, Office, etc) has brushing (all other lines are grayed out)
- Clicking a category name removes that line, or brings it back if already removed.

Other screenshots of interaction:

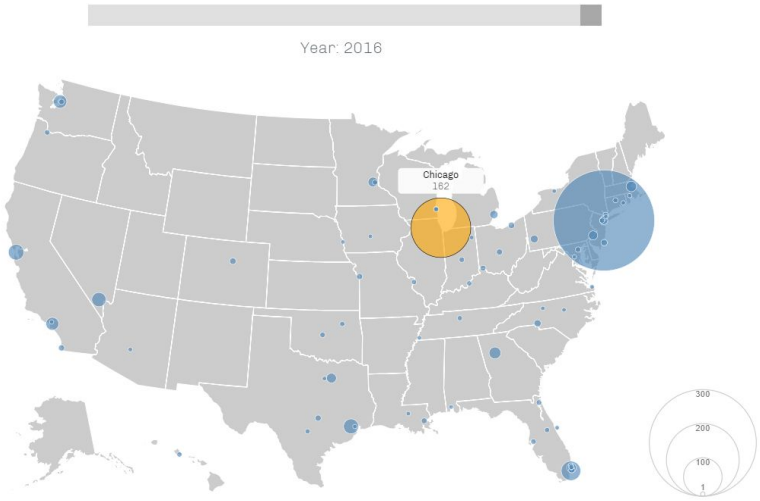


In the above example, “Office” is removed, and the mouse is hovering over “Residential.”



In the above example, “Office” and “Hotel” are removed, and the mouse is hovering over the chart at the x value corresponding to year 2005.

SYMBOL MAP WITH SLIDER

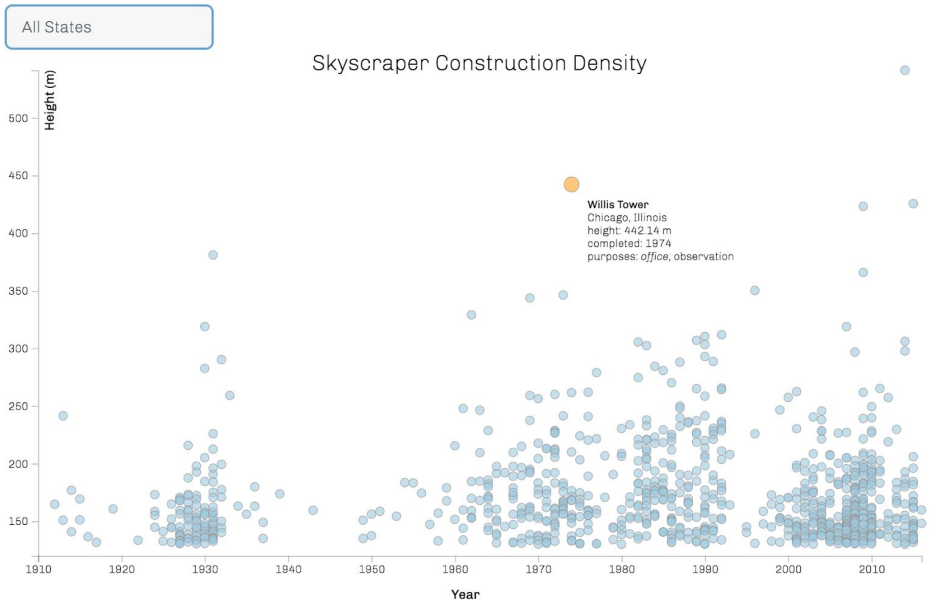


In the above example, the slider has been changed to Year 2016 and the mouse is hovering over the Chicago bubble.

Functionality:

- Slider at top allows you to adjust which year you would like to look at (transitions as you change the years, either moving forward or backwards)
- Hovering over a circle gives you city name and skyscraper count.

SCATTERPLOT

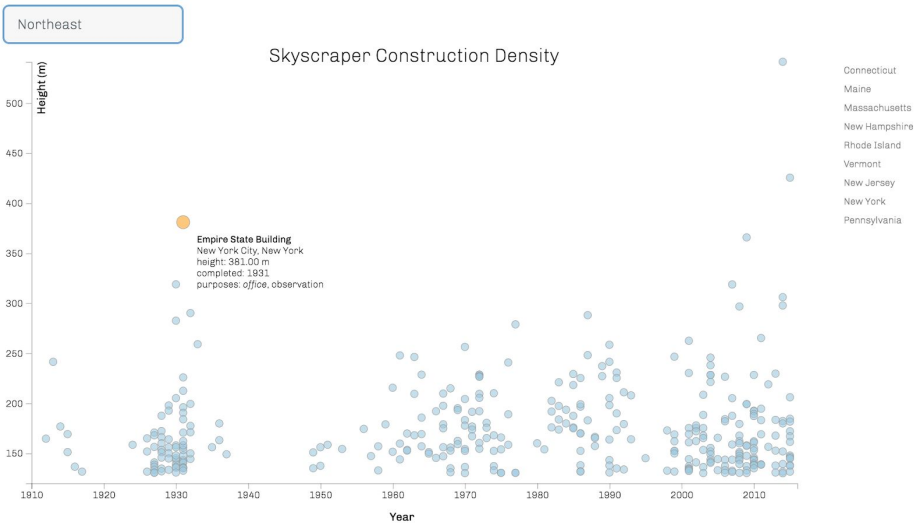


In the above example, "All States" is selected and the mouse is hovering over the point corresponding to the Willis Tower.

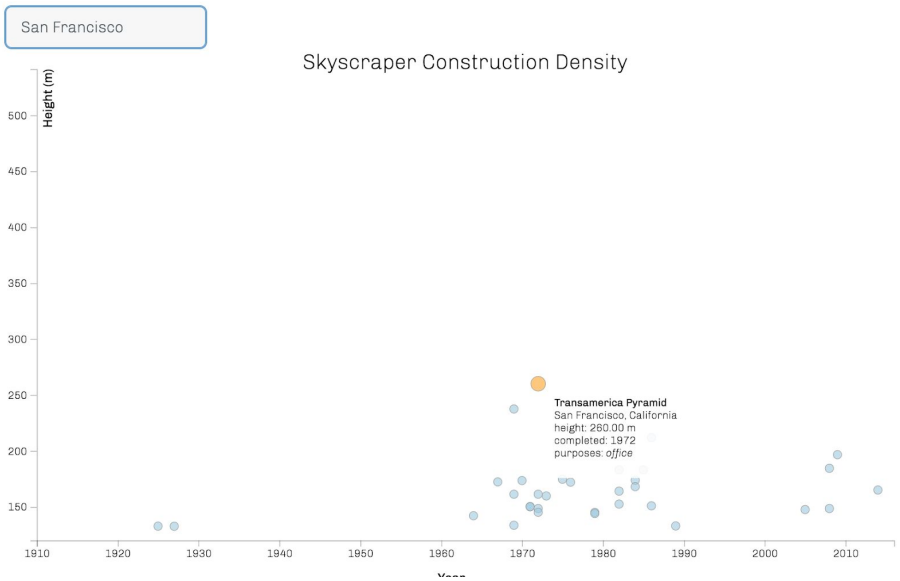
Functionality:

- Selecting the desired city/region from the drop down menu
- On selecting a different city, the dots transition and the tooltip information is updated
- Hovering over a dot reveals more information about the specific skyscraper: name, location, height, completion year, and purposes (all purposes listed, but main purpose *italicized*).

Other screenshots of interaction:

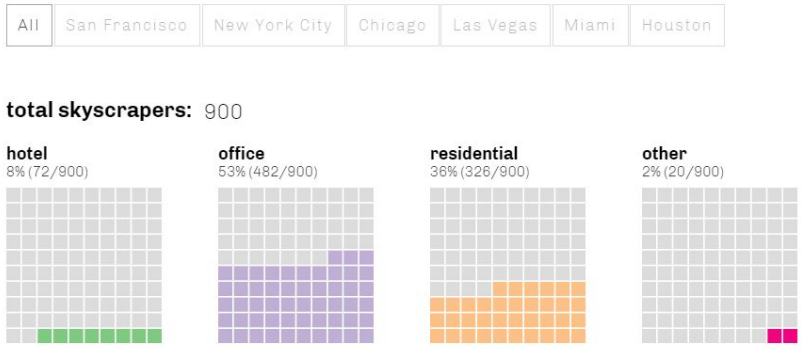


In the above example, the Northeast region is selected and the mouse is hovering over the point corresponding to the Empire State Building.



In the above example, San Francisco is selected and the mouse is hovering over the point corresponding to the Transamerica Pyramid.

ICON ARRAY

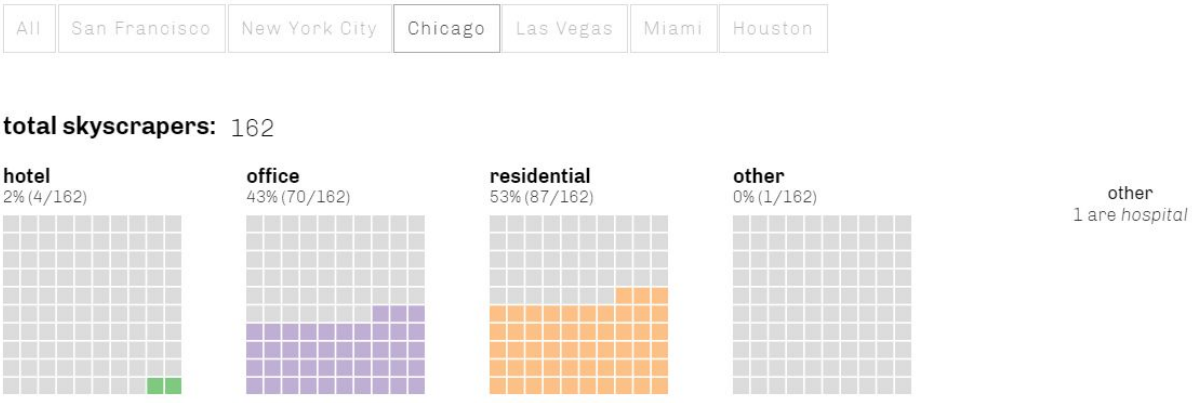


In the above example, "All" is clicked.

Functionality:

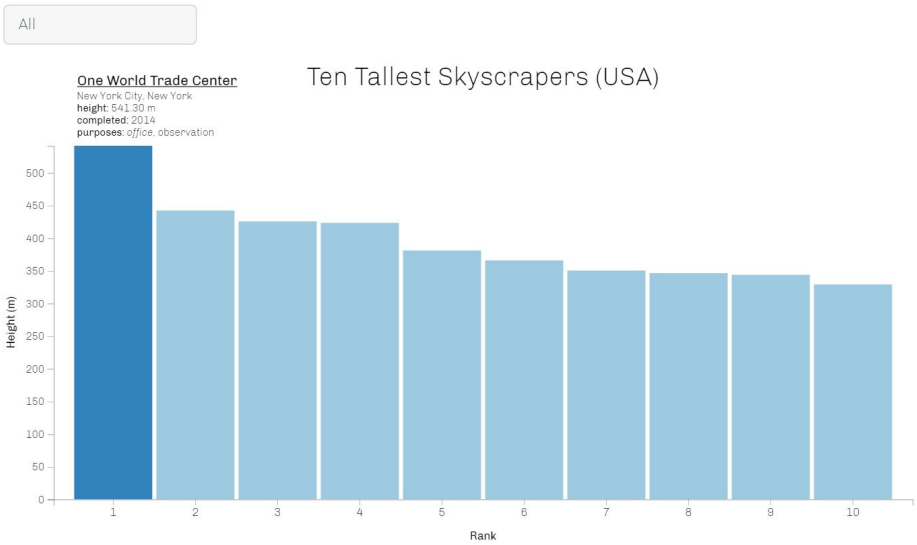
- Clicking on a city/All in the navigation bar at top changes the icon array values
- If at least 1 skyscraper in "other," breakdown information of what the other main purposes are given.

Other screenshots of interaction:



In the above example, "Chicago" is clicked and the mouse is hovering over "other."

BAR CHART

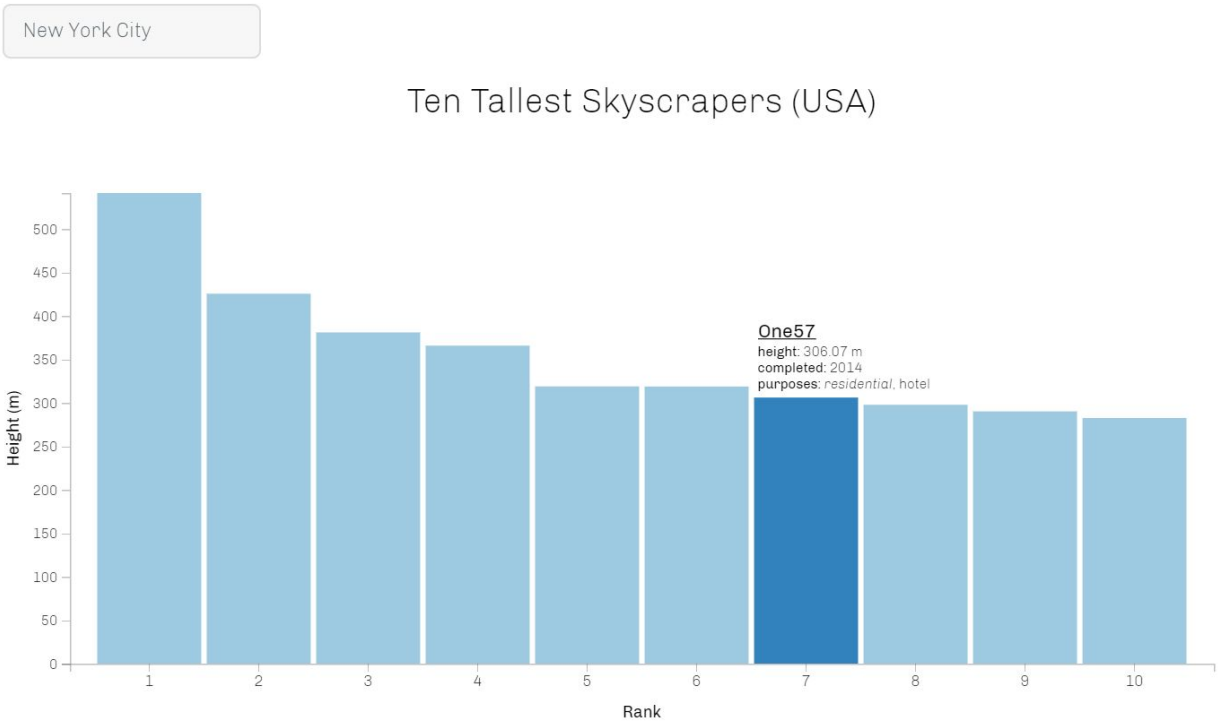


In the above example, "All" is selected and the mouse is hovering over the bar at Rank 1.

Functionality:

- Selecting the desired city from the drop down menu
- On selecting a different city, the bars transition and the tooltip information is updated
- Hovering over a bar reveals more information about the specific skyscraper: name, location (if in *"All"*), height, completion year, and purposes (all purposes listed, but main purpose *italicized*)

Other screenshots of interaction:



In the above example, "New York City" is selected from the drop down menu, and the mouse is hovering over the bar at Rank 7.

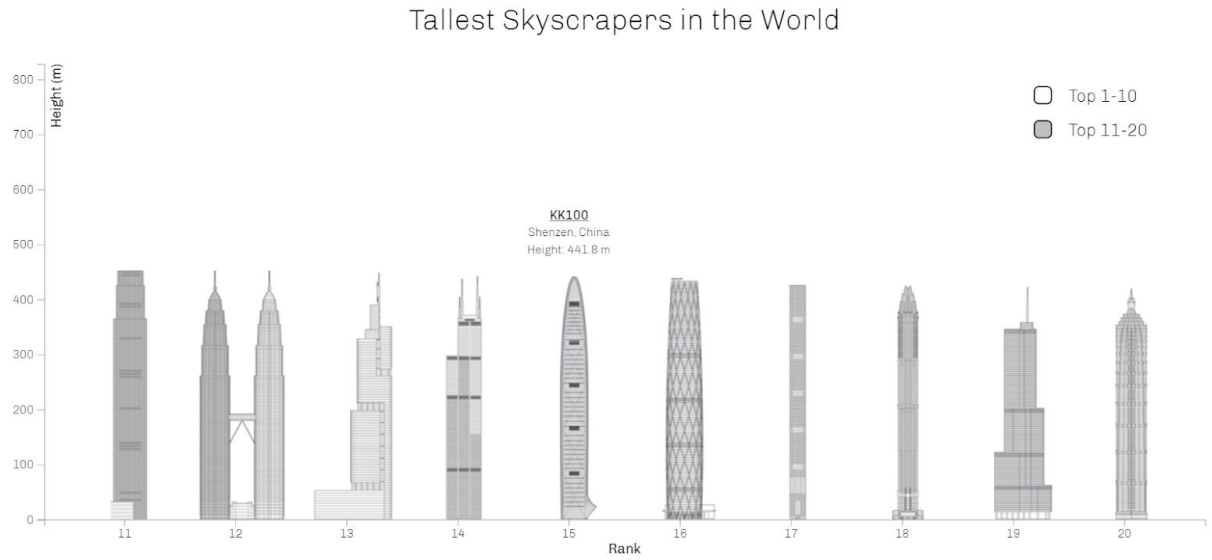
WORLD SKYSCRAPER COMPARISON

In the above example, "Top 1-10" is selected and the mouse is hovering over the image at Rank 4.

Functionality:

- Clicking on "Top 1-10" or "Top 11-20" changes the images and tooltips, and the box next to the selection turns gray (if the selection is different than the current selection)
- Hovering over an image gives tooltip with more information: name, location, height

Other screenshots of interaction:



In the above example, "Top 11-20" is selected and the mouse is hovering over the image at Rank 15.

Evaluation

What we learned from the data

We have found that there were periods in which skyscraper construction has slowed down (most likely due to military events) or bursted, and based on past trends, we do not expect skyscraper construction to slow down in the next few years. We also found the the east coast and west have a large difference in the number of skyscrapers constructed. More research into building/construction laws, population sizes, real estate, and the economy will be needed to do any further analysis.

The most common purpose for these skyscrapers are office buildings, then residential, then hotel, and then other purposes. However, it is not uncommon for these skyscrapers to have more than one purpose.

From the bar chart, we can see that the tallest skyscraper in the US is the One World Trade Center in New York City. It stands tall at 541.30m (1743.11ft). There is a noticeable difference between the first and second tallest building in the US. This Willis Tower in Chicago, Illinois is 88.16m (289.24ft) shorter than One World Center at 442.14m (1450.60ft). Overall, we find that most of the tallest skyscrapers in the US can be found in the New York City and Chicago, and these are the most populated in terms of skyscrapers, as well.

How does the tallest skyscraper in the US compared to the rest of the world? That would be incomparable. The tallest skyscraper in the world can be found in Dubai, UAE. At 828m (2716.54ft), the Burj Khalifa is is 287m taller than than the One World Trade Center. One World Trade Center ranks 6th overall in the world. We also find that China has a majority for the number of skyscrapers in the top 10.

How well do our visualizations work/improvements?

Our visualizations work pretty well, and we are proud of the visualizations we have achieved. We did deviate from our initial proposal quite a bit, but in a good way as we realized our initial designs were too simple, didn't work with the data, or we found better/new ideas.

Some of the improvements that could be made are better transitions. For example, in the icon array, the grids are removed and then redrawn each time a new city is selected rather than transitioning nicely (as in the inspiration example), but we ran out of time and could not figure it out.

It would also have been nice if we could have used the images of the skyscrapers in the bar chart as we did in the world visualization. However, because SkyscraperCenter did not have images for all skyscrapers, our visualization would have been incomplete. (Although, it also would've been a lot of work to manually scrape/clean/scale ~150 images)

Additionally, in the symbol map, it would be nice if there was less overlap among the East Coast cities. Perhaps if there was an easy way to implement a jitter. Implementing zoom would also be an option, but seemed unnecessary for this case given the overlap among cities was thankfully not too terrible. The bubbles are purposely made not fully opaque, given a white outline, and a bright orange color on hover over to alleviate the overlap problem as much as possible.

Furthermore, in the scatterplot it would have been nice to have a better way with dealing with the overlap of the dots. For now, we treated it in a similar way as in the symbol map by making the circles less opaque with an outline and making the circle on hover orange. We did try making the other circles that are not being hovered over fade to gray, but it was too distracting.