# Project Proposal Oct. 31

This summer, the NY state senate passed a bill to fine the hosts of certain Airbnb rentals in New York City. However, the governor of NY has yet to sign the bill. We’d like to present a website showing the context surrounding this controversial issue by showing both the prevalence of Airbnb rentals as well as the state of the housing rental market. We’ll show changes over time as well as trends by neighborhood or borough. We’ll also show the effects of illegal Airbnb rentals on lost hotel tax revenue and increased prices of standard rentals.

We’ll collect data from several online sources including insideairbnb.com and tomslee.net/airbnb-data. Both of these websites have scraped data from the Airbnb website directly at different time intervals. We’ll also download Zillow’s publicly available dataset of median rental prices in NYC. Finally, the Bureau of Labor Statistics provides a consumer price index, and we will incorporate their data from 2006 until now to adjust for inflation.

# Project Plan Nov. 7

## Goals and Tasks

* General goal: Explain the legal controversy and context surrounding a bill that has been passed in NYC to regulate Airbnb rentals
* Visualize Airbnb listings over time (allow users to explore several variables)

We want to give the users the tools to interact with our data on Airbnb postings. To this end, we’ll create a map with a slider that allows the user to show postings from a given map. They’ll also be able to select various attributes to color the posting dots (e.g., by price, neighborhood, etc). This will allow users to understand how Airbnb has grown in NYC in the last two years and also frame our discussion of its relationships with housing prices as well as lost tax revenue.

* Show relationships between Airbnb density and housing prices in NYC

Over the last several years, Airbnb listings in NYC have dramatically increased. In some areas, the cost of housing has increased as well. We want to show the rise in housing prices in NYC neighborhoods over time as a function of Airbnb density. We will use a line chart with each line representing a NYC neighborhood. The x-axis would be time (monthly from 2014 until 2016), and the y-axis would represent the percent change in Zillow Rental Index ([see their methodology here](http://www.zillow.com/research/zillow-rent-index-methodology-2393/)) from our starting date. The lines would each be colored by the density of Airbnb postings in the neighborhood (which we will determine based on the number of residential units, calculated from data provided by the PLUTO dataset from the NYC department of city planning). The user will see a general relationship between Airbnb densities and housing price increases across NYC. They’ll also be able to filter the neighborhoods and hover over lines to identify neighborhoods with high Airbnb densities and Zillow Rental Index increases. Finally, we’ll toggle between a view of general Airbnb density and illegal Airbnb density to show the user whether it is the illegal Airbnb listings in particular that are closely related to housing cost increases.

* Estimate lost tax revenue  
  Hotels in NYC must pay 3 types of hotel tax: state sales tax, city sales tax, and a base fee per room per night. Originally, Airbnb claimed that it was only a platform and, as such, individual hosts were responsible for reporting and paying tax on their Airbnb rental revenue. However, most hosts don’t report or pay those taxes. Using the nightly price for Airbnb’s over time and an estimate for how often listings are actually reserved, we can estimate the amount of tax revenue that each host would have to pay. Then, we can show how the lost tax revenue has changed over time and produce a final amount of total amount of revenue lost over the past 2 years.
* Show tax deals and other conflicts between Airbnb and other cities in the USA

This visualization will show context surrounding conflicts between cities and Airbnb regarding tax revenue. On a map of the United States, we’ll have markers of cities that have made deals with Airbnb as well as cities with ongoing lawsuits (e.g. San Francisco). The user will be able to hover over these cities to see details of resolutions or ongoing conflicts.

## Description of Data

We have collected data from several online sources including insideairbnb.com and tomslee.net/airbnb-data. Both of these websites have scraped data from the Airbnb website directly at different time intervals. For each listing, variables include latitude and longitude, price per night, description of the space, neighborhood, borough, host information, number of rooms, etc. We have also downloaded Zillow’s publicly available dataset of median housing rental prices in NYC, which breaks data down by neighborhood. We have collected data from the NYC Department of City Planning land use and geographic data on the number of housing units per neighborhood, from which we can calculate the density of Airbnb listings. Finally, the Bureau of Labor Statistics provides a consumer price index, and we will incorporate their data from 2006 until now to adjust for inflation.

We still need to collect data on tax deals Airbnb made with other cities, which we will be able to collect by researching news reports on such deals.

We also need to analyze the data in two ways:

1. Estimate which listings are illegal based on number of listings per host, host location, minimum nights to book, and type of property (and other variables if needed).
2. Estimate amount of lost tax revenue based on nightly prices and a model for how often Airbnb’s are booked (there are models for this online so we need to pick one that works).

## Project Timeline

* 11/7 Project Proposal due
* 11/14 Project Prototype 1 due
  + Deliverables:
    - Skeleton of website and one functioning visualization: Morgan and Michaela
    - Complete necessary assumptions and calculations to produce analysis - Alena
    - Cleaned data - Nicasia
    - Draft 1 of copy - Morgan
    - All team members make Github accounts and become familiar with using it
* 11/28 Project Prototype 2 due
  + Deliverables:
    - All visualizations functioning and partially interactive
    - Draft 2 of copy
    - Final website layout
* 12/12 Final Project due
  + Deliverables:
    - All visualizations functioning and fully interactive
    - Copy finalized

## Feature List

### Must-Have

* Map with airbnb listings as nodes in New York, showing change over time
* Line chart of percent increase in housing prices in certain neighborhoods
  + Highlight neighborhoods with some threshold density of Airbnb listings: #airbnbs / #residential units
* Change in illegal listings before and after publishing a report about Airbnb in New York City (Nov 2015)
* Bar chart of tax dollars “lost” over time due to Airbnb
  + City tax dollars “lost”
  + State tax dollars “lost”

### Good-to-Have

* Cities across the US where Airbnb has made tax deals (with tooltips with details of the tax deals)

## Team Roles:

Within each iteration of webpage design, each team member will be responsible for taking the lead and the following roles and making sure all related tasks are carried out:

* Target - Alena
* Data Wrangling - Nicasia
* Design - Morgan
* Implement - Michaela
* Evaluate - Alena

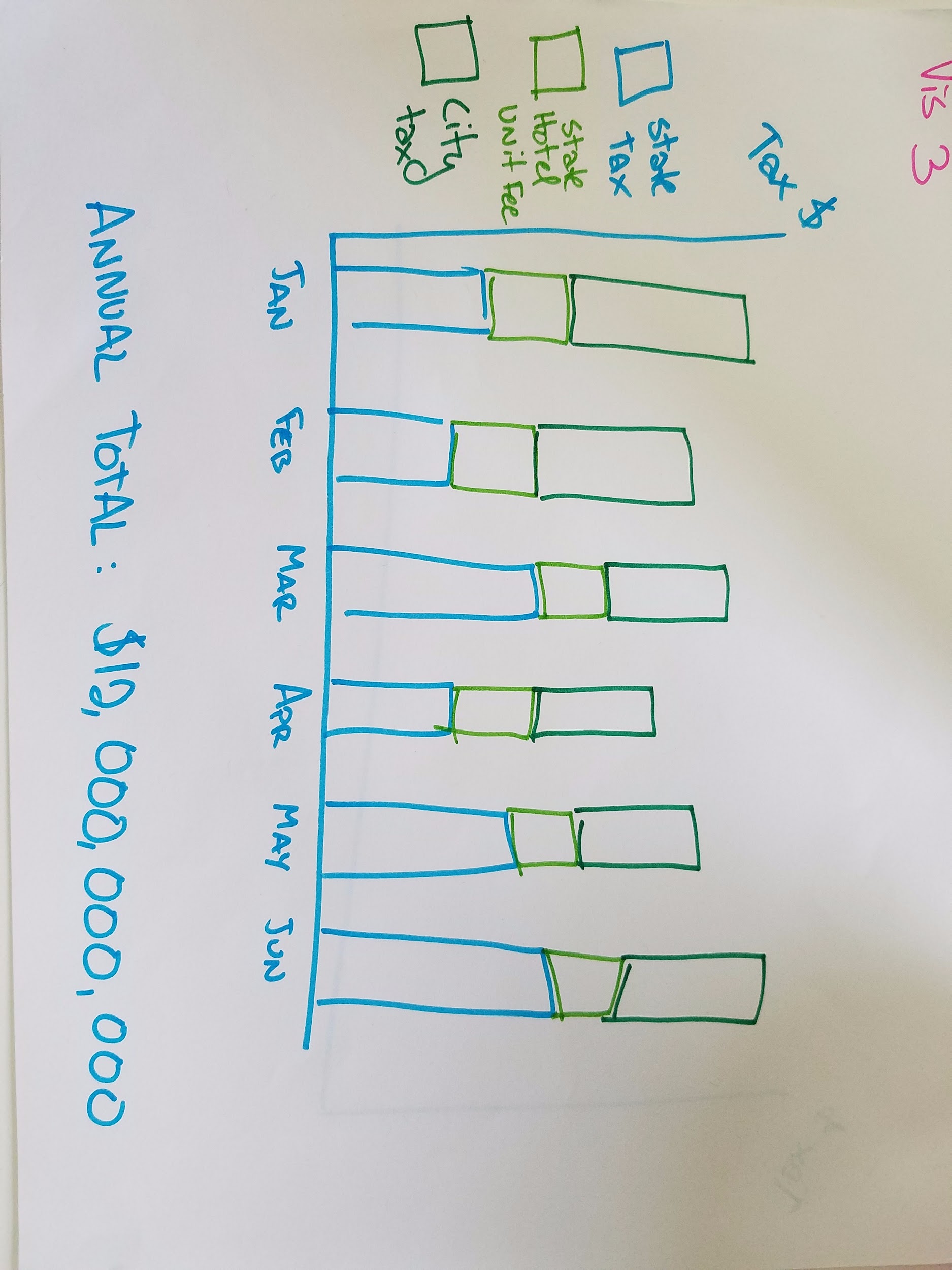
Presentation in Studio November 8

After presenting our project proposal in Studio, we received the following feedback:

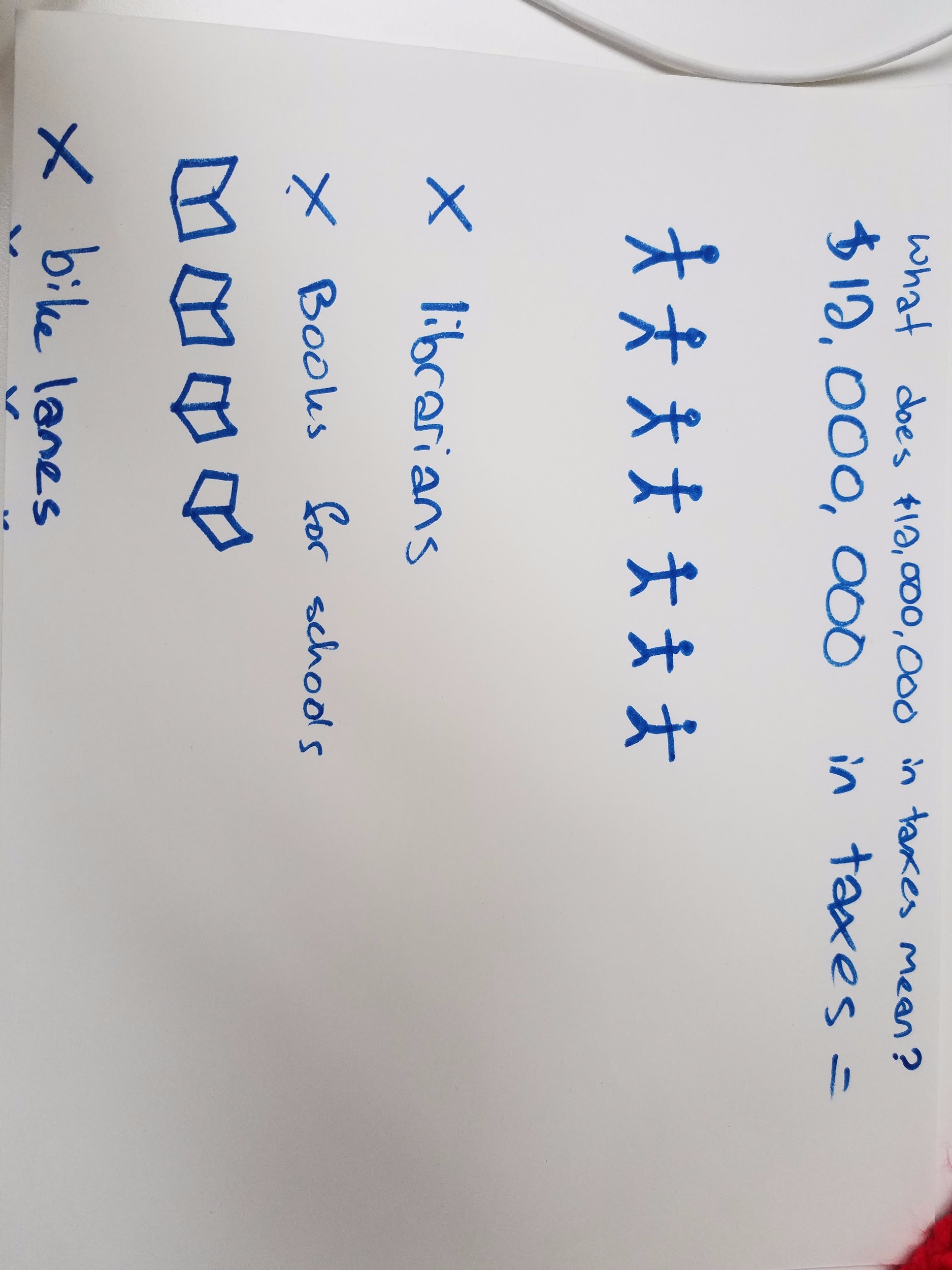
* If we’re presenting this as the two sides of a debate, that should be reflected in the structure of the project, i.e. explicitly presenting the cases of both sides.
* We need to make sure we stay focused with our scope, and everything connects together.

# Creative Visualizations from Lecture November 10

From earlier, we decided that our stacked bar chart showing lost tax dollars each month may be too dry and difficult for users to interpret. In order to help individuals better visualize and understand the amount of tax dollars lost.



One way we might orient the user to the amount of money lost may be to show how the money translates to the cost of public services offered by NYC. For example:

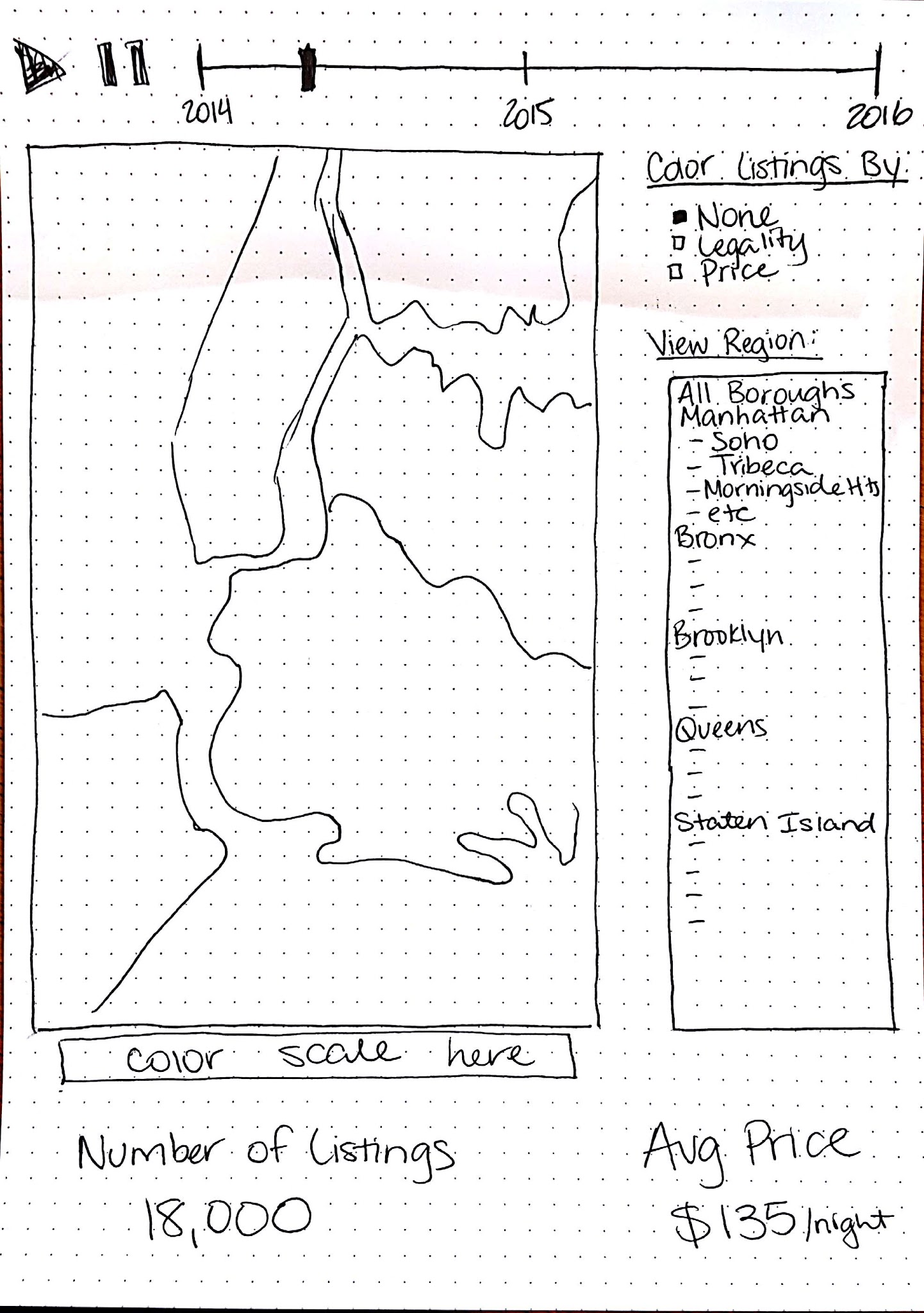


After designing this visualization, we decided that our drawing may be somewhat misleading because Airbnb taxes clearly would be divided and spent on various budgetary expenditures in both the city and state government. Therefore, it may make more sense instead to create a bar graph representing NYC’s budget broken up into different divisions (e.g., parks and rec) and then add a bar representing the lost tax revenue from Airbnb, so that users can compare this value to the various NYC budgets.

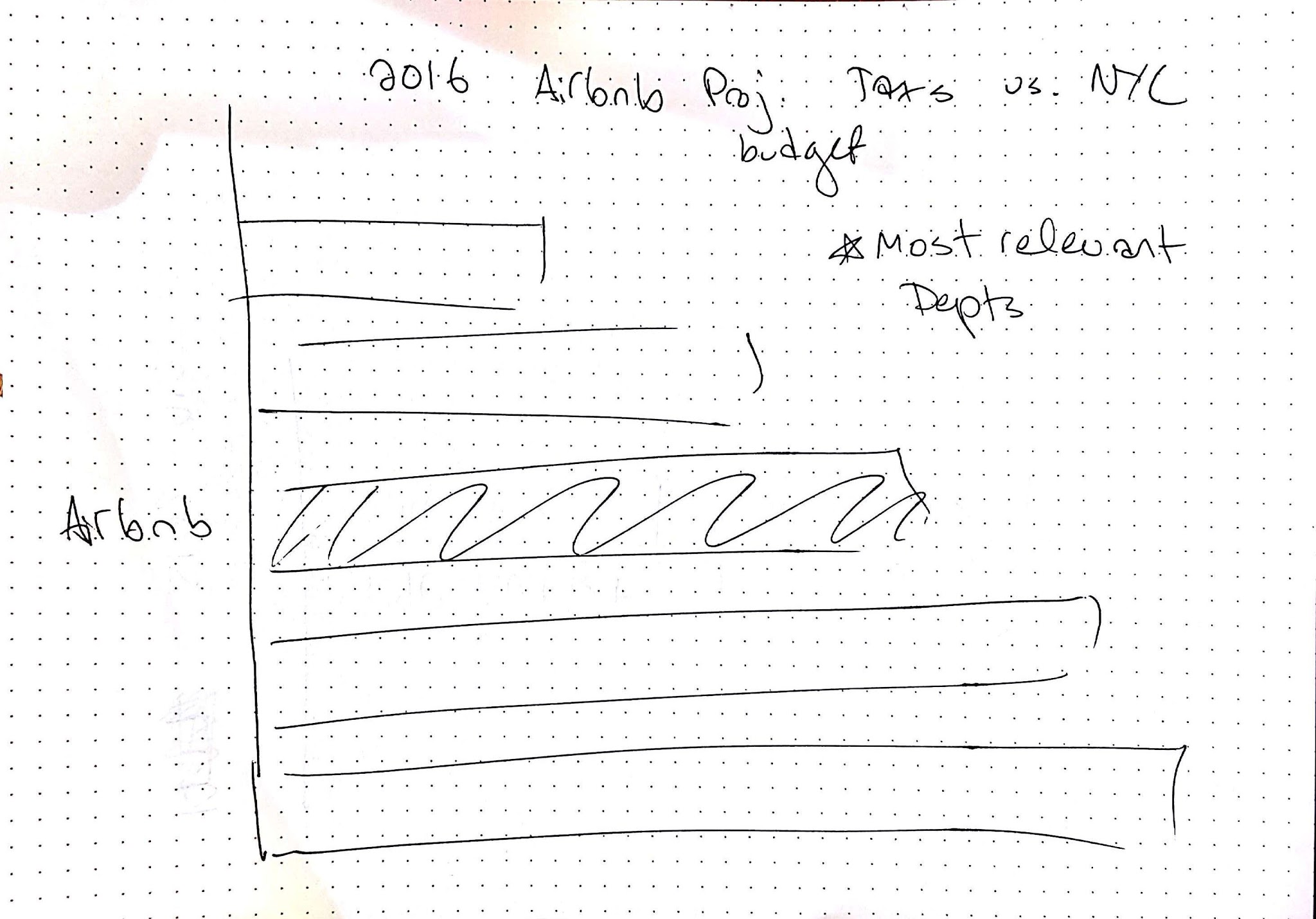
Project Meeting: Prototype V1 November 12

* Skeleton of website looks good, but we should amend the navigation links to stick to the top of the page.
* Morgan will take a stab at designing a header image
* How to figure out legality and taxes:
  + Alena figured out what the laws are and put together a plan to calculate it, and wrote code snippets to do so.
  + Assumptions/Estimates doc
  + Assumption: Supply and demand of Airbnb listings will not change based on imposition of taxes
* Issue for prototype of visualization: So much data that the web page is crashing before it loads
* Reconsidering visualizations:
  + Swap out second map for Sankey diagram showing breakdown - found a plugin that produces them. This will help readers understand what makes something illegal, and the quantity of illegal listings.
  + Consider swapping bar chart of taxes over time with bar chart putting NYC tax revenue in context of other
  + For first visualization:
    - Cut priceband slider
    - Give context to issues coming up in the introduction so the map makes sense - Basically, you shouldn’t be running a hotel out of your house
    - Neighborhood data: Dropdown to zoom to neighborhood or click on neighborhood, or hover over neighborhood to show name, cluster by neighborhood

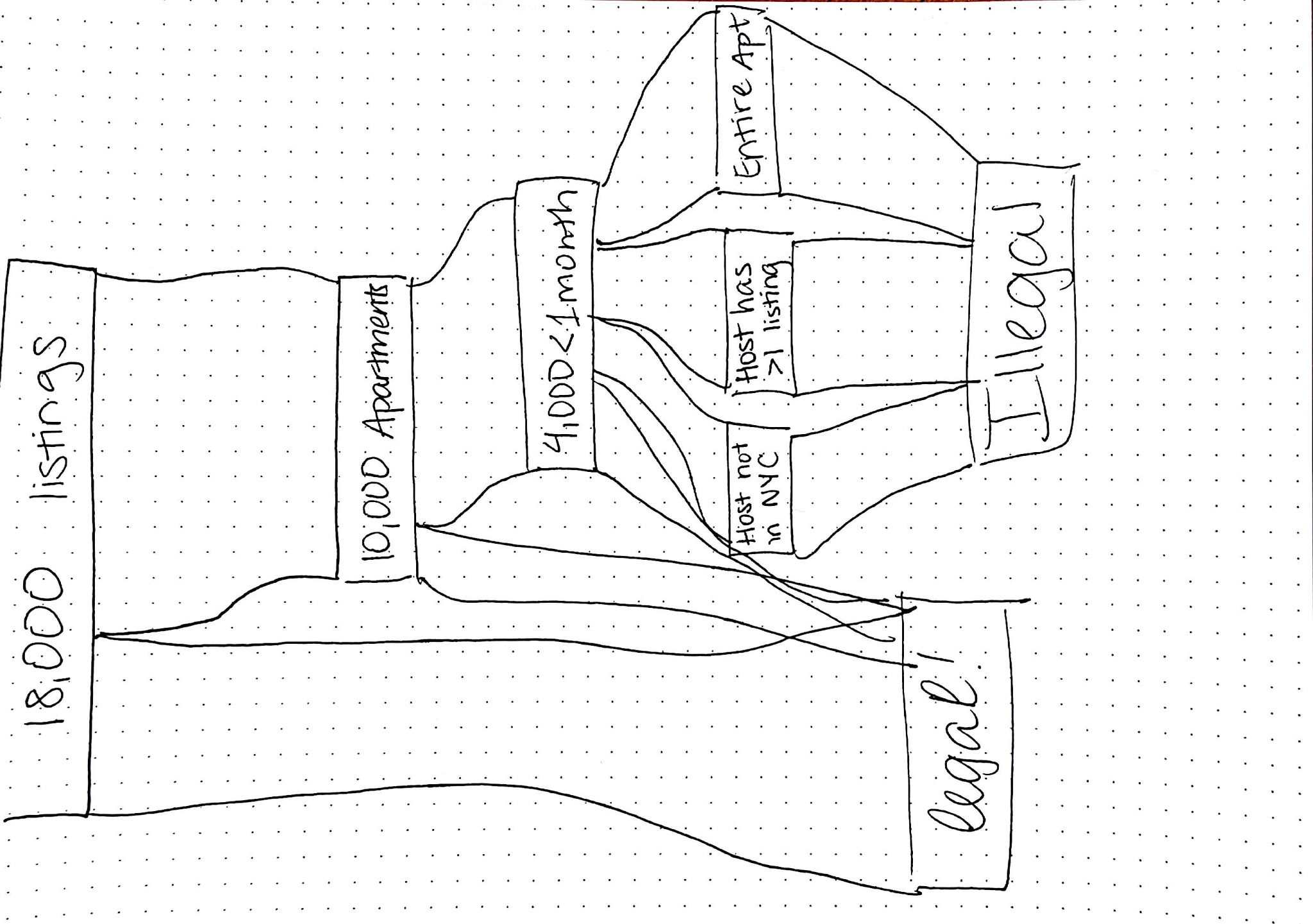
First visualization (revised)



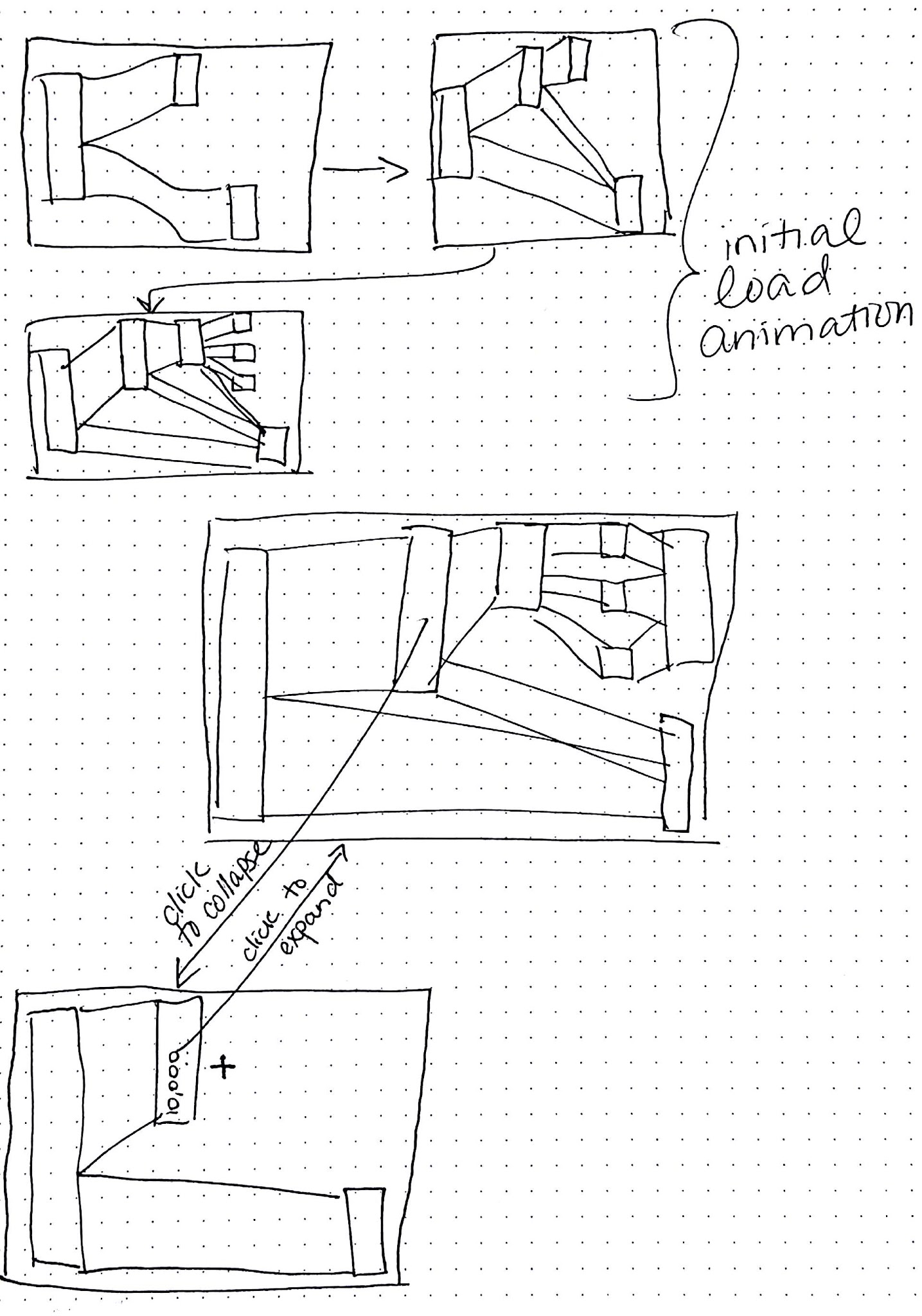
Contextualized Taxes Bar Chart



Sankey Diagram



Sankey Storyboard



Tasks for upcoming week:

Morgan:

* Header Image
* Comparison tax bar graph
* Template for line graph (everything except coloring by density)

Alena:

* Sankey diagram with dummy data

Nicasia:

* Finish data cleaning for one data set - Monday
* Finish getting data by next meeting

Michaela:

* Keep working on Vis 1 - Monday

# Legality & Tax Methodology

**Number of illegal listings**

“Since 2010, it has been illegal in New York to rent out a whole apartment for fewer than 30 days.”

<http://www.nytimes.com/2016/10/22/technology/new-york-passes-law-airbnb.html>

“In [New York City](http://www.nolo.com/legal-encyclopedia/overview-airbnb-law-new-york-city.html), for example, residential property located in a multiple residential dwelling unit, such as an apartment building, must be used for “permanent resident purposes.” This means that the property must be occupied by the same person or family for 30 or more consecutive days. It's illegal to have paying guests for less than 30 days--unless, of course, the property is a licensed hotel, bed-and-breakfast, or other similar business. However, there is an important exception: It's not illegal to rent a room in New York City if you occupy your home or apartment at the same time and all parts of the dwelling are available to the paying guest.”

<http://www.nolo.com/legal-encyclopedia/legal-restrictions-renting-your-home-airbnb-other-rental-services.html>

Illegal if:

* Multiple residential dwelling unit (apartment)
* AND occupied by paying guests for < 30 days

Legal if:

* If owner is still there
* AND all parts of dwelling are available to the guest

Based on available data, illegal if:

|  |
| --- |
| (d.property\_type == "Apartment") && (d.minimum\_nights < 30) && (d.room\_type == "Entire home/apt") |

This is literally the definition

|  |
| --- |
| (d.property\_type == "Apartment") && (d.minimum\_nights < 30) && (d.host\_listings\_count > 1) |

Host could conceivably be renting out the rooms of the apt to different people, but if they’re renting out multiple apartments then they’re clearly not living in that one

|  |
| --- |
| (d.property\_type == "Apartment") && (d.minimum\_nights < 30) && (!d.host\_location.includes("New York")) && (!d.host\_location.includes("NY")) && (d.host\_location != "US") |

If host doesn’t live in NYC then they’re clearly not in this apt

*Property names from July 2 2016 file*

*We can’t really know whether all parts of the dwelling are available to the guest. It’s probably a safe assumption that they’re not (who wants a rando in their bedroom?) but we can’t really account for that.*

**Lost tax revenue**

1. **What taxes are owed**
   1. Total tax on an NYC hotel room: 14.75% + $3.50/room/night (this is assuming the room costs at least $40/night which I think is a fine assumption for us)
      1. NYC sales tax: 4.5%
      2. NY State sales tax: 4%
      3. MCTD state sales & use tax: 0.375%
         1. Money to MTA (hooray for subway)
      4. NYC Hotel Room Occupancy Taxes:
         1. People are exempt from paying this if ([source](http://www1.nyc.gov/site/finance/taxes/business-hotel-room-occupancy-tax.page))
            1. Rooms only rented for < 14 days. **Aka if d.maximum\_nights < 14**
            2. OR Rooms only rented once or twice in a fiscal year.
         2. NYC Hotel Occupancy Tax Rate: 5.875%
         3. NYC Hotel Room Occupancy Tax of $2.00/room /night applies to all rooms charging $40/night or more.
            1. This is for every room. So if you have a suite you have to pay more than $2. ([source](http://www1.nyc.gov/site/finance/taxes/business-hotel-room-occupancy-tax.page))
      5. NYS Javits expansion fund charge of $1.50/room/night applies to all rooms, effective 4/1/05.
         1. Also per room (i.e. suite consideration)
         2. I believe has same exemptions as above
         3. >tfw NYC expands the javits center and makes tourists pay for it
2. **Occupancy rate**
   1. Average NYC Occupancy Rate: 62.36% (9/24/16 [source](https://www.mashvisor.com/blog/best-cities-airbnb-occupancy-rates/))
   2. Alt: 80% (1/12/16 [source](https://smartasset.com/mortgage/where-do-airbnb-hosts-make-the-most-money))
   3. Inside Airbnb claims 33% estimated occupancy (based on 10/1/16 data [source](http://insideairbnb.com/new-york-city/#)). [Methodology](http://insideairbnb.com/about.html#disclaimers):
      1. Assume review rate of 50% to convert # reviews to # of bookings
      2. Average length of stay is based on published information (Airbnb says it’s 6.4 nights in NYC - [source](http://blog.airbnb.com/airbnbs-economic-impact-nyc-community/))
         1. If published info is not available, they use 3 nights per booking as average length of stay
         2. If listing has higher minimum # nights then average length, they use the min # for that listing
         3. Occupancy rate is capped at 70%

**What we can do:**

1. Adapt the Inside Airbnb occupancy method to calculate a # nights booked/month for each listing. Cap the # nights booked at 70% of the month.

|  |
| --- |
| var avgLength = 6.4;  var maxNights = 30 \* .7;  var stayLength = (d.minimum\_nights > avgLength) ? d.minimum\_nights : avgLength;  d.nights\_booked = d.reviews\_per\_month \* 2 \* stayLength;  d.nights\_booked = (d.nights\_booked < maxNights) ? d.nights\_booked : maxNights; |

1. Use the nightly price and the nights booked/month to calculate monthly revenue:

|  |
| --- |
| d.month\_revenue = d.nights\_booked \* d.price; |

1. Determine if this listing has to pay the NYC Hotel Occupancy Tax:  
   (I don’t think we can figure out if it’s booked just 1 or 2 times per fiscal year, but we can see if the max stay is under 2 weeks)

|  |
| --- |
| d.hotel\_tax\_applies = (d.maximum\_nights < 14) ? false : true; |

1. Determine number of rooms in the listing  
   We have # bedrooms for each listing, but I think it’s safe to assume that if you are renting an entire home you also have access to some kind of living room so...

|  |
| --- |
| d.num\_rooms = (d.room\_type == "Entire home/apt") ? (1 + d.bedrooms) : d.bedrooms; |

1. Calculate taxes!

|  |
| --- |
| d.tax.NYC\_sales = .045 \* d.month\_revenue;  d.tax.state\_sales = .04 \* d.month\_revenue;  d.tax.mctd = .00375 \* d.month\_revenue;  d.tax.hotel\_occupancy = (d.hotel\_tax\_applies) ? (.05875 \* d.month\_revenue + 2 \* d.num\_rooms \* d.nights\_booked) : 0;  d.tax.javits = (d.hotel\_tax\_applies) ? (1.5 \* d.num\_rooms \* d.nights\_booked) : 0; |