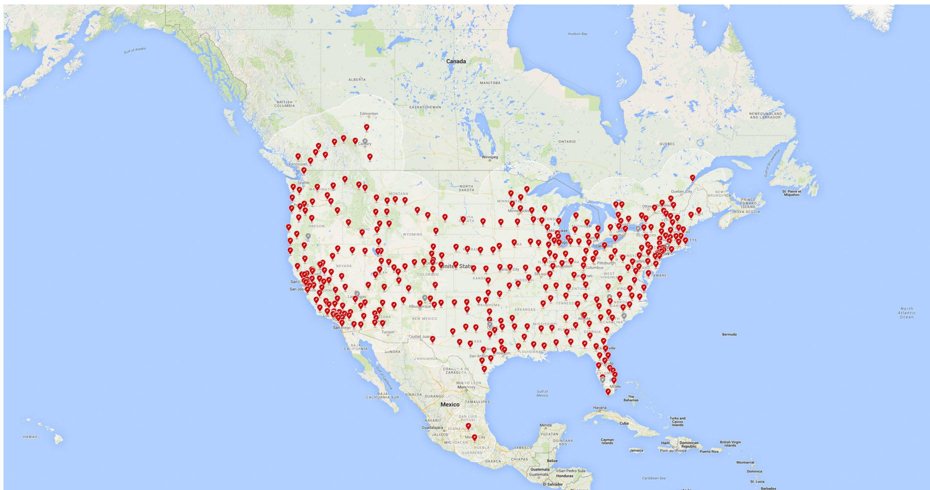
Introduction

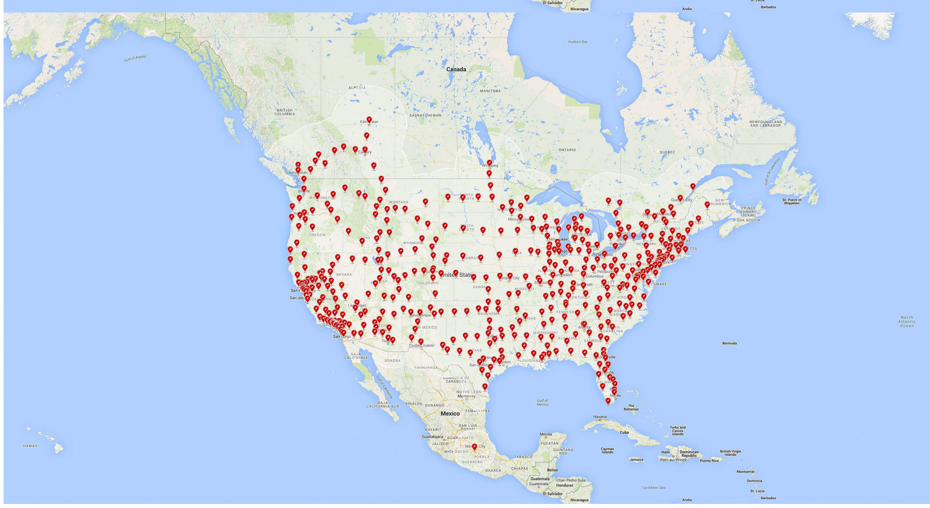
Tesla Super Charger Network Exploratory Data Analysis

Introduction

With the recent unveiling of Tesla's Model 3 and pre-orders approaching 400,000, the internet has been buzzing with Tesla discussions and analysis. One of Tesla's key differentiators from other mass market Electric Vehicals (EVs) is its Super Charger (SC) network that provides 170 miles of range in 30 minutes source. (https://www.teslamotors.com/supercharger) With Elon Musk stating plans to double the size of the SC network by the end of 2017, a large amount of planning, resources and investment are being allocated to this network expansion.

Tesla 2016 Planned North America Network Growth





source (https://www.teslamotors.com/supercharger)

Project Overview

The goal of this project is to build an intelligent decision support system to provide recommendations for where Tesla should expand their Super Charger (SC) network to maximize its effectiveness in creating a robust charging infrastructure for EVs in North America. In order to develop this system, an exploration of the current network and associated data is needed to better understand the criteria and attributes Tesla may be using to develop their network, and to provide insight into which parameters provide the best decision power for this network.

Assumptions

This analysis brings many assumptions along with it, which are documented here to highlight and clarify potential (and highly probable) sources of error. Statements throughout the analysis that contain inherent assumptions listed here are denoted by @# where # represents the numbered assumption in this list.

1. The population for each SC is the sum population of cities within a 73km radius of the SC's GPS coordinates from the availible dataset of 5209 cities. Assumption is this will be an adequate estimation for all SC locations in the dataset.
2. Network connections are weighted based on combined population of the connecting nodes. Assumption is this is an adequate proxy represenation of the strength of a connection between SCs for the supercharger network, although it is obvious that the density of Tesla owners in each city will likely have an impact on the strength of connection. Inherent in this assumption is that the amount of Tesla owners in a city is directly proportional to the population of that city. Although this is clearly not the case, it is assumed to follow close enough for the purposes of this analysis.
3. SC nodes are connected if they are within the maximum base range of a Tesla model 3 (215miles - 346km source (https://www.teslamotors.com/model3)) as determined by the driving distance computed by google directions API between SC nodes.
4. Assumption that Tesla's underlying network growth mechanism is likely about maximizing the breadth and market penetration fo the network
5. The number of direction steps retrieved from the google directions API is used as a proxy for the complexity of travel between 2 SCs and that less steps between 2 SCs likely

suggest their close proximity to major highways, as it makes sense that more distance travelled on a major highway should reflect in less steps required in the google directions. This assumption has not been confirmed, and the actual implementation of the google directions API step count may not adequately model this interpretation
6. Assumed desirable characteristics based on personal intuition

EDA Plan

With so many variables to combine and explore, a plan is needed to organize the analysis. The analysis is broken into 3 main sections that represent the 3 "meta" datasets requiring exploration:

1. Static Network Characteristics - The current state of the network and its structure
2. Dynamic Network Characteristics - The growth of the network over time and space
3. Supercharger Data - Exploration of the amenities of each Super Charger
4. Network Growth Utility - Exploration of the likely desirable traits an ideal network would have

Initial Setup

Utility Function Imports and Data Preperation

The following scripts setup the required packages used in this analysis and prepares the network data for EDA consumption.

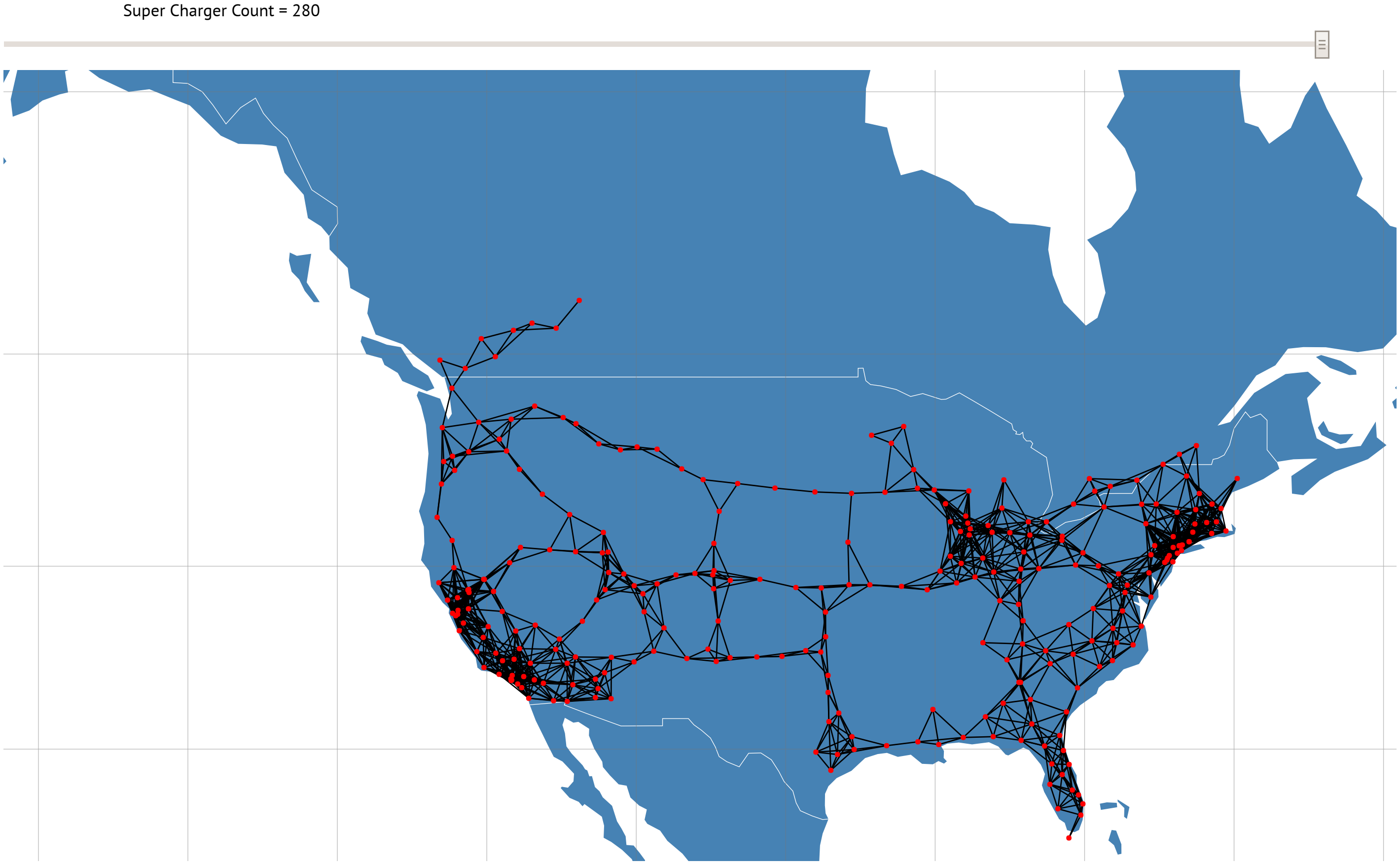
```
Attaching package: 'dplyr'
The following objects are masked from 'package:stats':
    filter, lag
The following objects are masked from 'package:base':
    intersect, setdiff, setequal, union
```

Network Graphic

The following graphic displays the Super Charger network with existing nodes and connections. The graphic can be manipulated to display its connection state at each point of network growth using the Super Charger Count slider.

Note - this graphic is dependent on the d3.js file sc_network.js that needs to be copied to a folder called "tesla_sc_network" in beakers root serving directory (\src\main\web). The d3.js can be download from the github repo here (https://github.com/cole-maclean/MAI-CN/tree/master/Final%20Project). Alternatively, the graphic can be viewed directly here (http://cole-maclean.github.io/MAI-CN/)

Super Charger Count = 280



Data Model

Data Model

Data Sources

The collected data has been structure into a NetworkX Graph, which is made up of nested dictionaries. There are 4 main datasources that were used to build the data structure:

1. Teslarati's Super Charger Map (http://www.teslarati.com/interactive-tesla-supercharger-map/) - Used as the source of existing Tesla Superchargers with GPS coordinates
2. Google Maps Directions API (https://developers.google.com/maps/documentation/directions/) - Used to abtain driving distances between each SC @3
3. Tesla Info (https://www.teslamotors.com/findus/location/supercharger/chamberybarberazsupercharger) - Information about amenities at a specific SC (ie. Chargers, Restaurants, WiFi, etc.)
4. Population Data (http://ezlocal.com/blog/post/Top-5000-US-Cities-by-Population.aspx) - List of cities and corresponding population.

Both Nodes and Edges have nested data dictionaries containing their respective datum from the above datasets, with example structures displayed below.

Notes

1. Nodes are keyed on the geohash of the GPS coordinates given for each SC from teslaraties dataset
2. The weight of each edge is the normalized aggregated population of the connected nodes - (pop_node_1 + pop_node_2)/(total_population_dataset)@1@2 where total_population_dataset = 200,087,175

```
{'City': 'Dayton',
 'Country': 'USA',
 'Elev': '258',
 'GPS': '39.85675, -84.276458',
 'GPS_lon_lat': [-84.276458, 39.85675],
 'Open Date': '2014-11-11 00:00:00',
 'SC_data': {'Chargers': ['8'],
  'Charging': ['8 Supercharger stalls, available 24/7'],
  'Lodging': ['Comfort Inn and Suites'],
  'Restaurants': ['Chipotle', "McDonald's", "Steak 'n Shake (open 24 hrs)"],
  'Restrooms': ['Meijer',
   'Chipotle',
   "McDonald's",
   "Steak 'n Shake (open 24 hrs)"],
  'WiFi': ["McDonald's", 'Comfort Inn and Suites']},
 'SC_index': '133',
 'Stalls': 8.0,
 'State': 'OH',
 'Status': 'Open',
 'Street Address': '9200 N Main St',
 'Supercharger': 'Dayton, OH',
 'Tesla': 'http://www.teslamotors.com/findus/location/supercharger/daytonsupercharger',
 'Thread': 'http://supercharge.info/service/supercharge/discuss?siteId=241',
 'Zip': '45445',
 'city_state': 'Dayton_OH',
 'geohash': 'dph4dpxy2hwn',
 'lat': 39.85675,
 'lon': -84.276458,
 'population': 7719014.0}
```

```
{'distance': 302588,
 'first_node': '99',
 'lon_lat_1': [-97.33313983306289, 37.60878003202379],
 'lon_lat_2': [-99.31914193555713, 38.90054295770824],
 'second_node': '164',
 'steps': 10,
 'weight': 0.009135588794741984}
```

Static Network Analysis

Summary Statistics

```
Stat     Value
avg_edge_distance 215238.35968379446

clustering_coef 0.6207007325759731
avg_short_path 9.482616487455196
edge_count 1518
node_count 280
avg_edge_weight 0.048406751667838
avg_degree 10.842857142857143
```

Summary Statistics

Reviewing the summary statistics, the Tesla SC network currently has 280 SCs in North America, connected by an average distance of 215kms or 62% of the Model 3's base range. With 1,518 total connections and an average of 11 connections per SC, each connection bridges nearly 5% of the total availible population and an average shortest path between any 2 SCs being 10 nodes. With 62% of clustering triangles being connected (ie clustering coef = 0.62), high density clustering of the current network can be seen near the North American coasts, and sparser clustering existing through the central mainland.

Static Network Node Characteristics

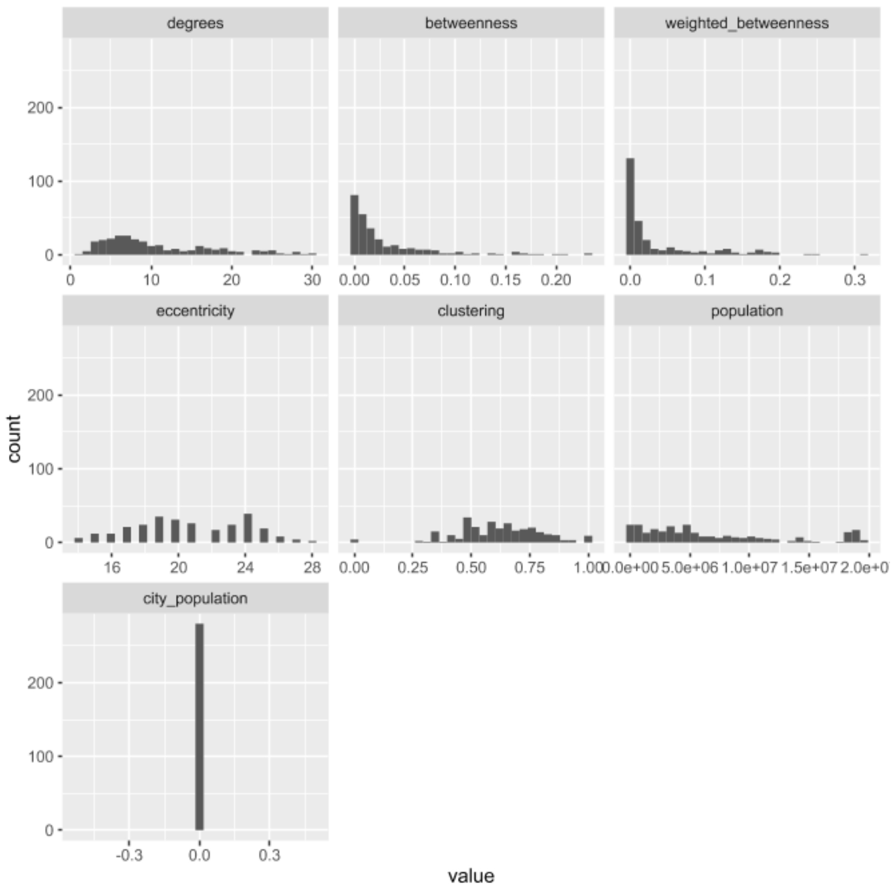
Network Characteristic Distributions

Using NetworkX's built-in algorthms, various network characteristic metrics are computed and plotted. The first set of characteristics are node dependent, and the second set are edge dependant.

	Index	Value

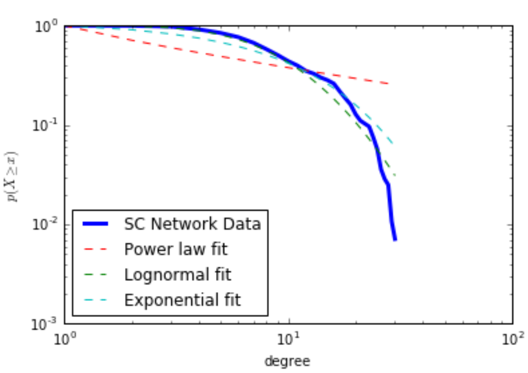
0	degrees	25
1	betweenness	0.134572818992836
2	weighted_betweenness	0.168882958149609
3	eccentricity	25.05000000000001
4	clustering	0.884438405797101

```
No id variables; using all as measure variables
```



Degree Distribution

The degree distribution of the SC network appears to be a long-tailed normal distribution or log-normal distribution. It does not appear to follow the typical "scale-free" distributions of many real world networks such as internet connections or social networks. The degree distribution is explored further below. Code adapted from source (http://nbviewer.jupyter.org/github/psinger/notebooks/blob/master/powerlaw_basics.ipynb)



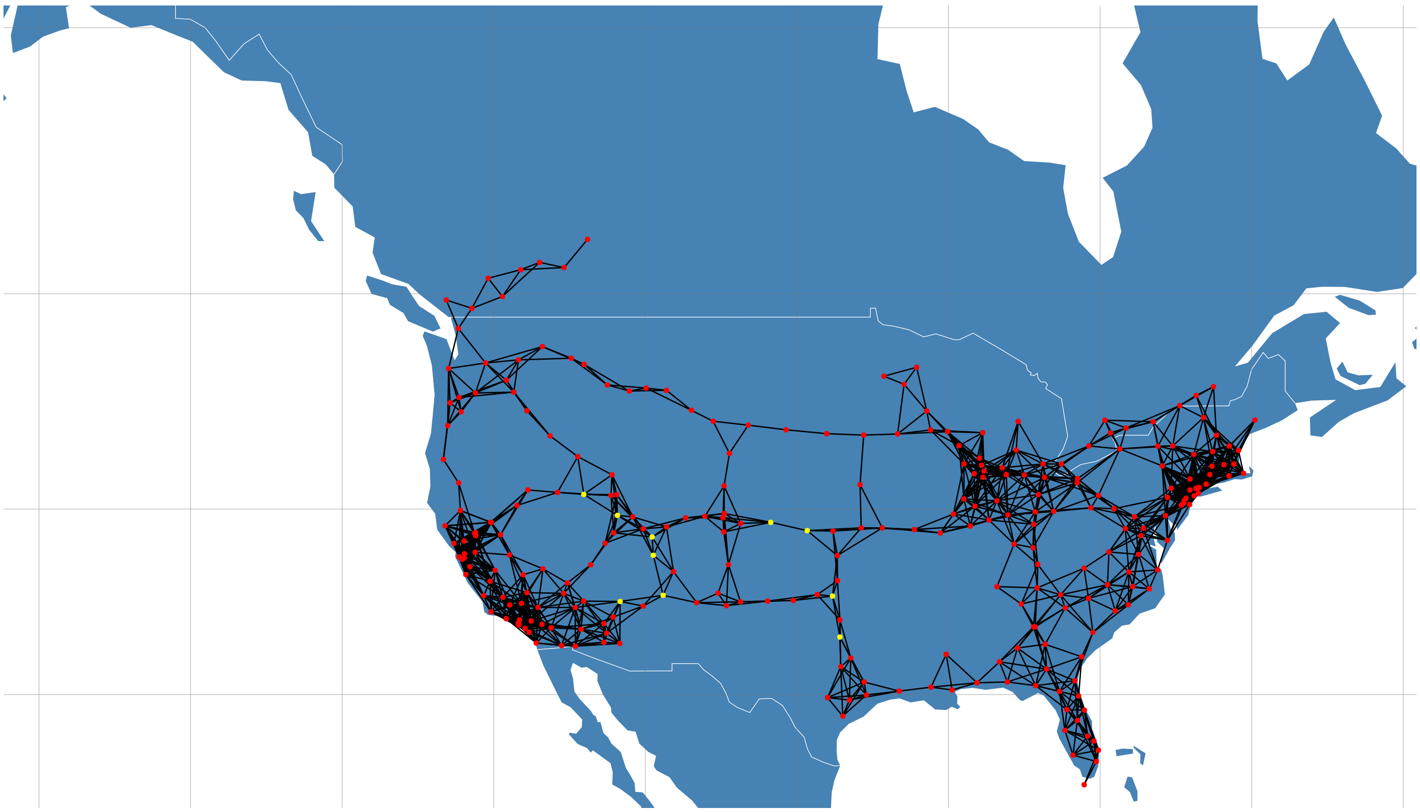
SC Network Degree Distribution

Consistent with observations of the degree distribution histogram, the powerlaw fit does not accurately model the degree distribution of the SC network. The lack of fit to a power law distribution makes intuitive sense, as networks that are modelled by power law distributions often have a "preferential attachement" or "the rich get richer" interpretation, where the most popular nodes (ie nodes with the most connections) are more likely to receive more connections to it source. (https://en.wikipedia.org/wiki/Scale-free_network) This reasoning runs counter to the likely impetus of the SC network, where the goal is to spread the networks breadth as opposed to the development of hubs that are generally attributed to power law degree distribution networks. @4

The degree distribution is better modelled as lognormal, suggesting a growth mechanism that varies significantly from the "preferential attachement" models of complex networks. One such growth mechanism that is said to produce lognormal degree distributions is Gibrat's Law, where the proportional rate of growth of a firm is independent of its absolute size source. (https://en.wikipedia.org/wiki/Gibrat%27s_law) This growth mechanism may be more appropriate for modelling the SC network, as it makes intuitive sense that the growth in number of connections a SC node has is independant of its current connection count. Stated otherwise, the mechanism of growth for the SC network does not depend on the current degree of SC nodes. SC, A, is connected to a new SC, B, because B increases the breadth and market penetration of the network and is independent of the number of connections currently attached to A. @4

SC Network Betweeness

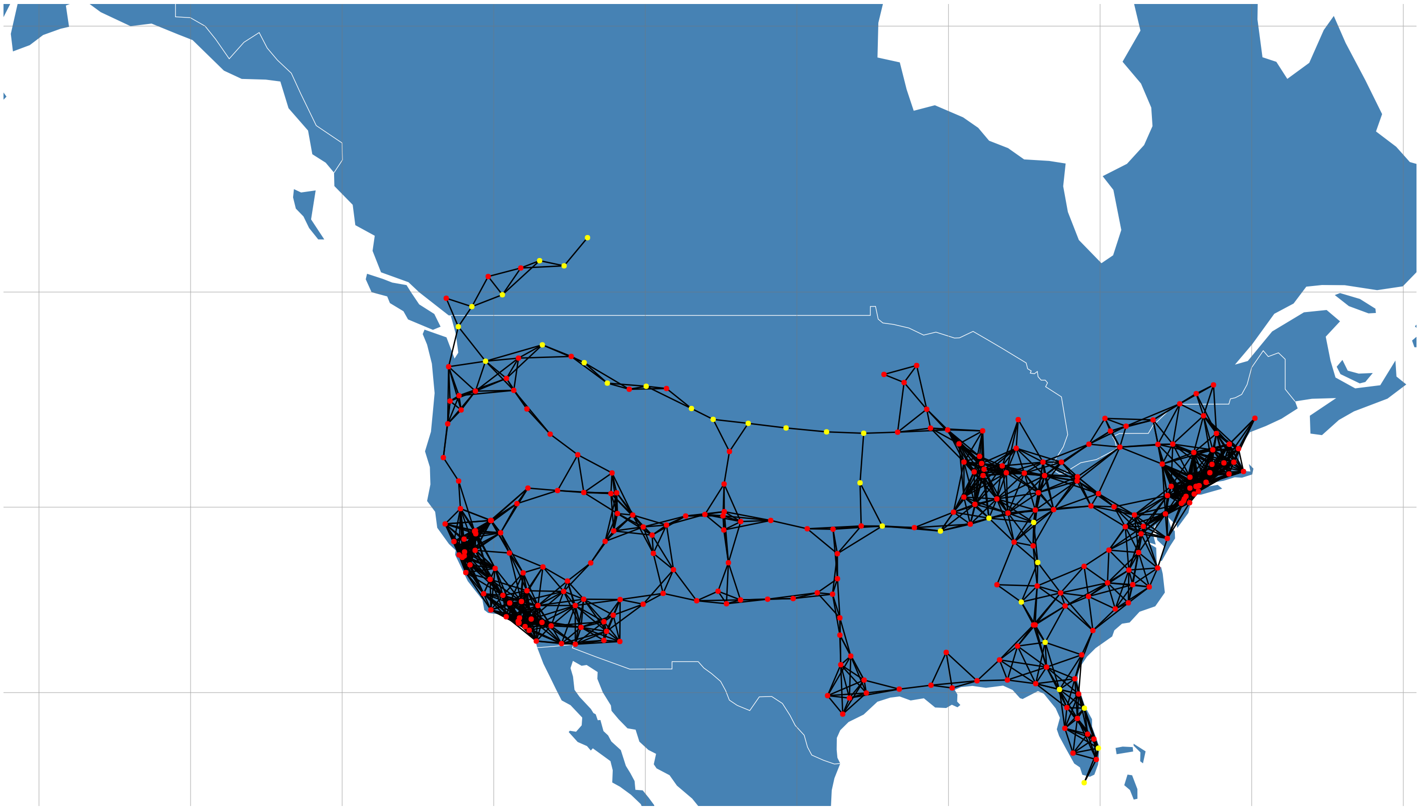
Unlike the degree distribution, the betweeness distribution of the SC networks does appear to follow more closely a powerlaw distribution. This means that mosts SC nodes have very few other nodes with their shortest path passing through them to reach the rest of the network, but a few nodes have a large amount of nodes with their shortest paths passing through them., making these nodes critical for network connectivity. The distribution is even more dramatically scewed to lower betweeness ratios when weighted by the connected populations, but the long tail remains with 1 SC node having a betweeness coefficient connecting 30% of the total potential population.



The top 10 nodes with the highest betweeness are highlighted in yellow in the above graphic. The majority of these high betweeness nodes appear to be the SCs connecting the California west coast to the east coast communities.

SC Network Eccentricity

The eccentricity distribution of the SC network seems to be a double humped normal distribution, with few nodes having mid-ranged eccentricity of about 23. In fact, a node representing an eccentricity of exactly 23 is missing from the entire network. Eccentricity is the measurement of the maximum path length that exists for each node. Reviewing the network, this double hump distribution makes sense, due to the 2 edge communities in the California and New York regions, along with the more central upper mid west community, causing 2 peaks in the distribution of paths between these communities. The maximum eccentricity of a network is know as its diameter, with a networks periphiery being the list of nodes with eccentricity equal to its diameter. The periphery of the current SC network goes from the tip of Florida in Marathon to just north of my home town of Calgary, in Red Deer, Alberta with a total of 28 SCs along the way.

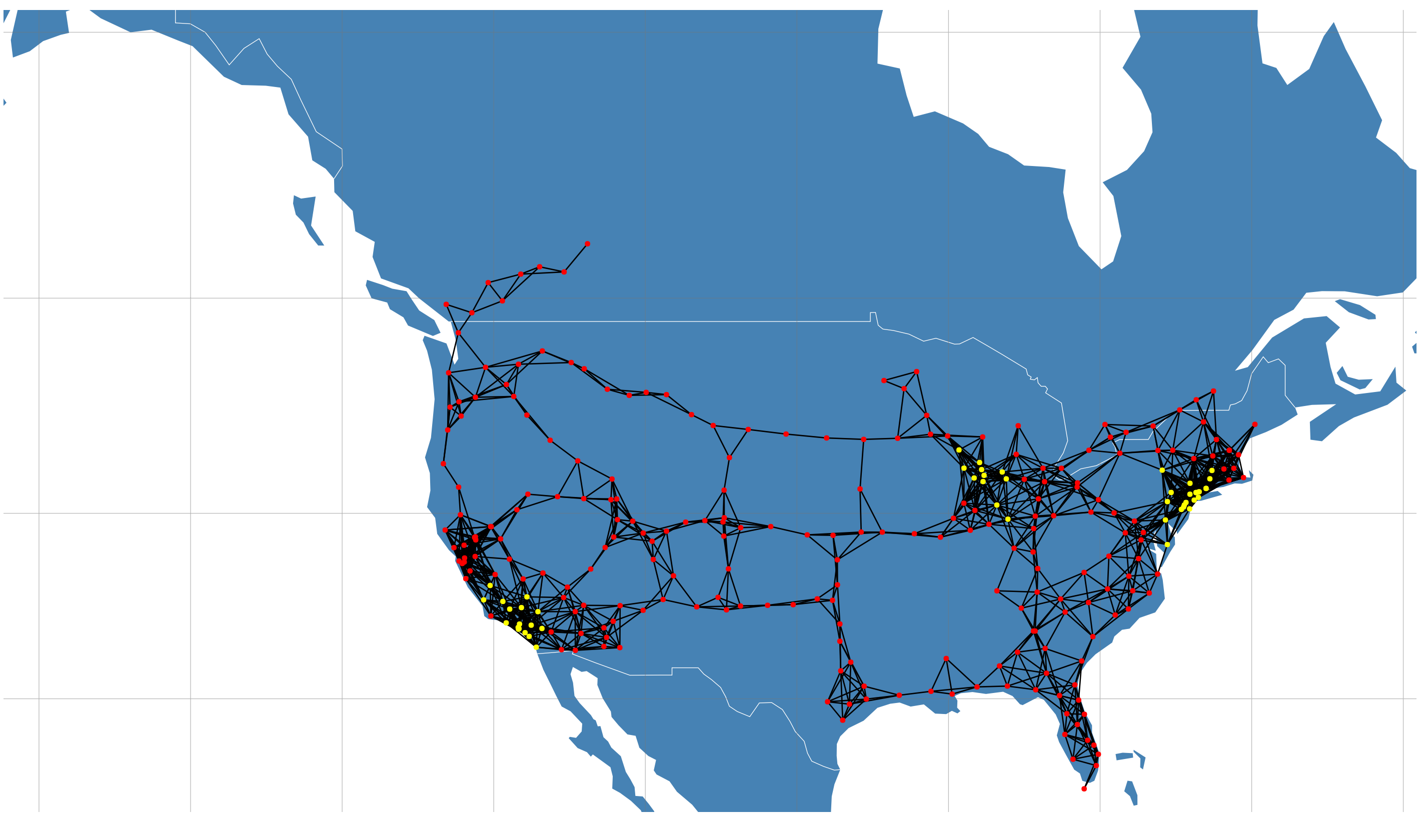


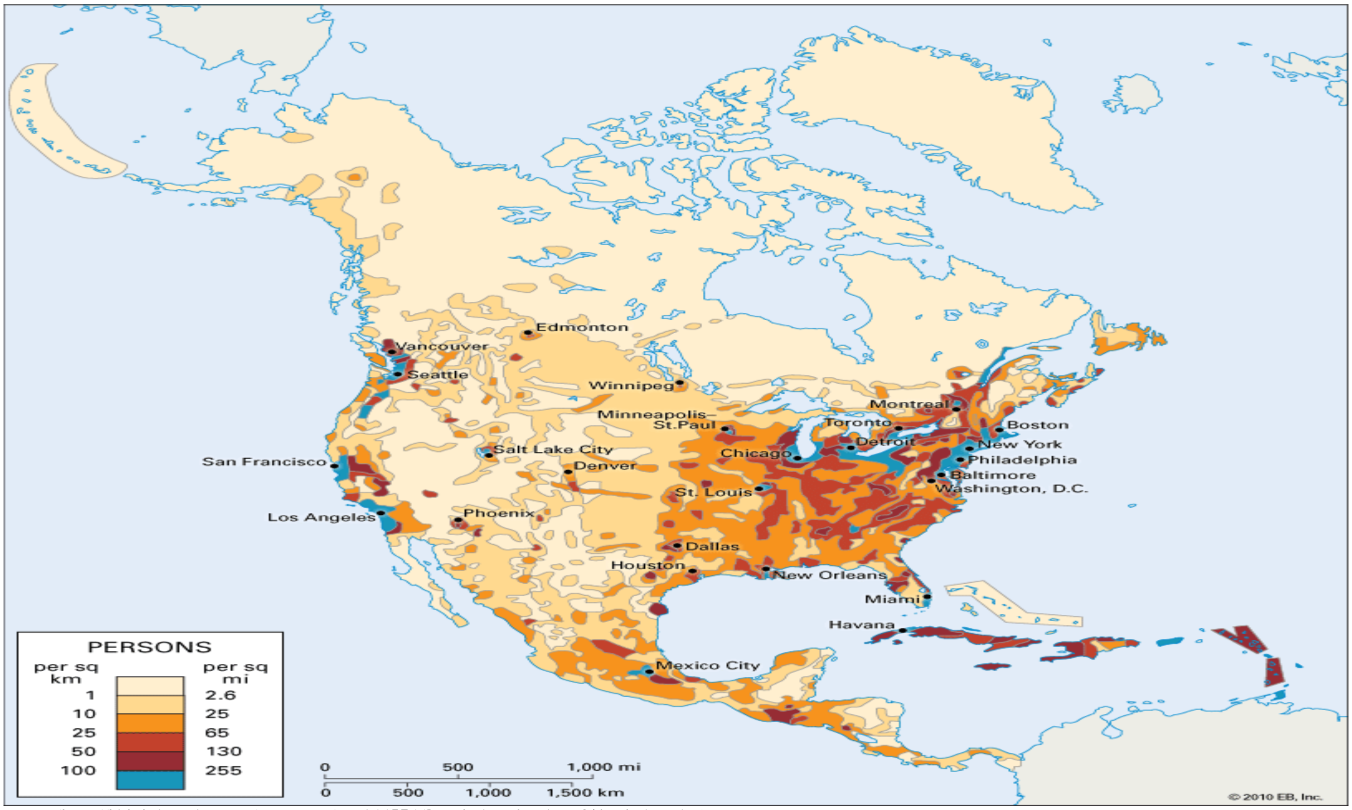
SC Network Clustering

The distribution of clustering for the network is normally distributed, which is to be expected as older SC locations (ie. California coast) become matured and filled in while the network continues to expand its edges (ie Canadian cities) resulting in lower nodes with low cluster coefficients.

SC Populations

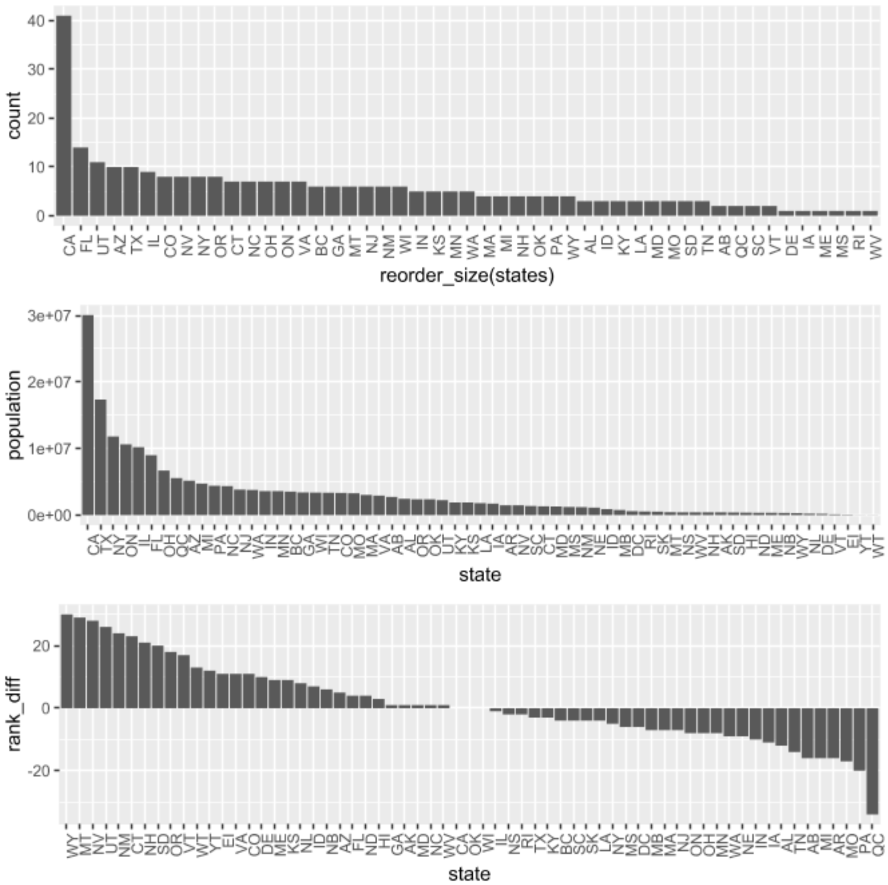
The population distribution of the network is fairly flat, with a few high population outliers from the high population density areas highlighted in the below graphic. Reviewing North America population density maps gives support to the theory that SC clusters are formed around high density geographic locations.





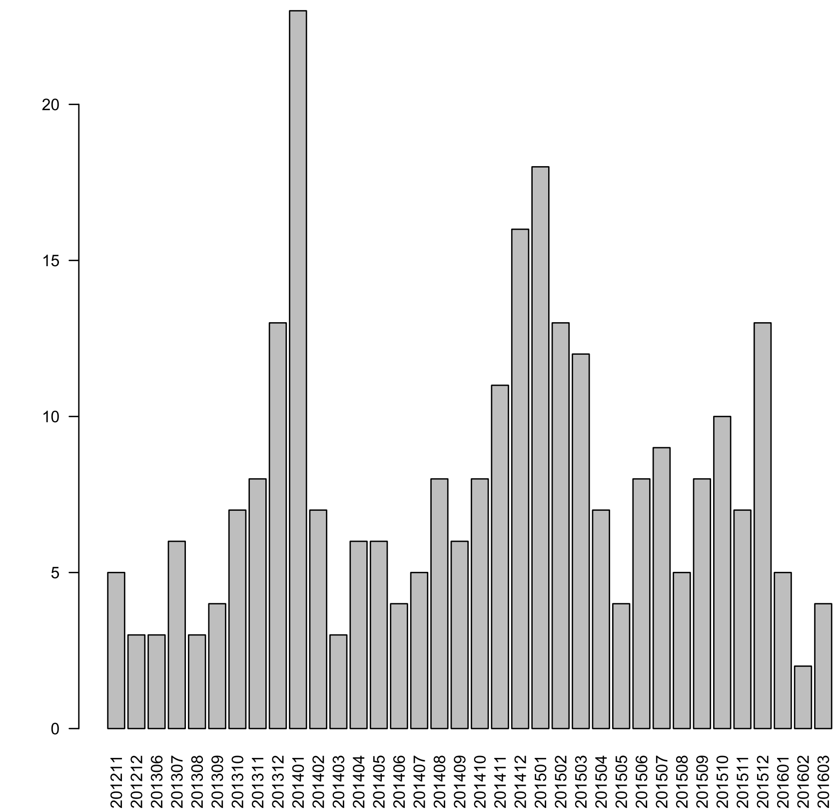
source (http://kids.britannica.com/comptons/art-166536/Population-density-of-North-America)

```
   state  rank_diff
56    WY         30
46    MT         29
34    NV         28
28    UT         26
39    NM         24
   state  rank_diff
9     MI        -16
33    AR        -16
21    MO        -17
10    PA        -20
7     QC        -34
```



States SC Counts vs Population

As was expected, California is by far the most represented state in the network, which makes sense both because of California being the most populace state in the dataset and Tesla being headquarter there. In fact, California is one of the only states to have its SC representation rank match its population rank. Other states have a larger disparity between their SC count and population ranking, with Wyoming being the most over represented state and Quebec being the most under represented province. This analysis illustrates that the network expansion considers much more then just population densities to decide the optimum SC locations, which is a fairly trivial statement, but it is interesting to see how far from the ideal flat 0 line the real distribution actually is.



SCs per Month

The distribution of SCs vs month seems to be cyclical, which may have to do with construction or permitting cycles, but the peaks of the cycles seems to be decreasing, which is opposite of the anticipated distribution. I expected to see an acceleration in the opening of SCs, but that does not seem to currently be the case. There is also a gap in new SC openings between January 2013 and May 2013.

Static Network Edge Characteristics

```
weights                    0.095064
distance              335541.900000
connection_age_diff      229.000000
steps                     27.000000
dtype: float64
```

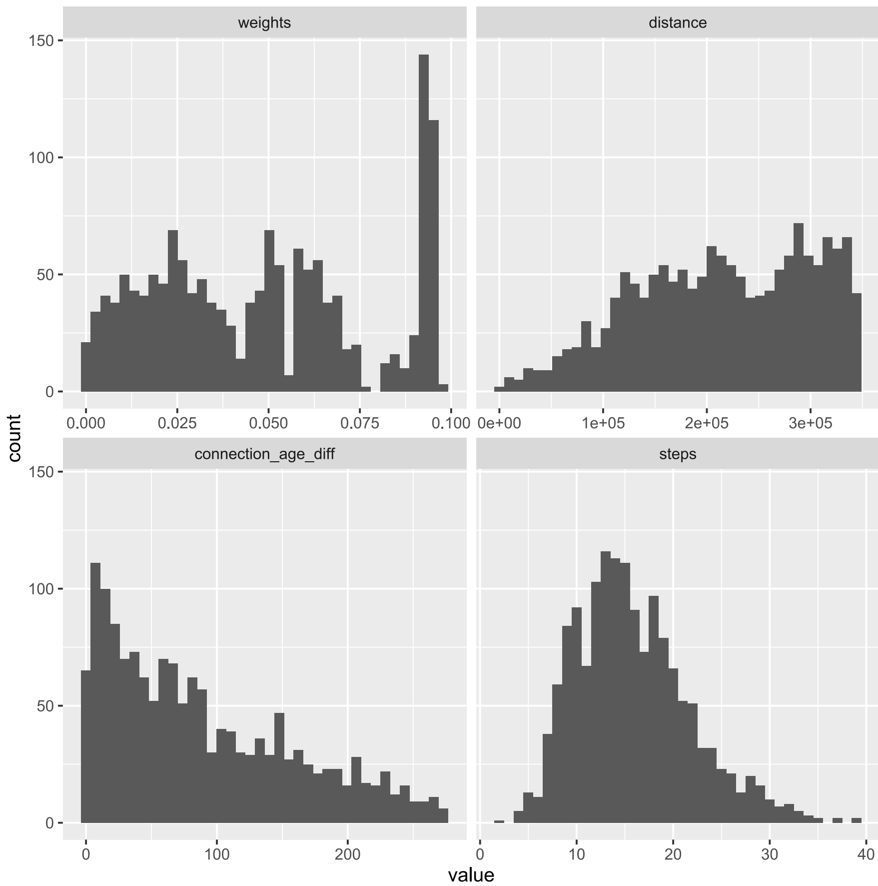
Index	weights	distance	connection_age_diff	steps
50%	0.0484	217722.5000	71.0000	15.0
75%	0.0685	289700.0000	138.0000	19.0
count	1518.0000	1518.0000	1518.0000	1518.0
max	0.0982	345838.0000	274.0000	39.0
mean	0.0484	215238.3597	88.3294	15.8
min	0.0003	1444.0000	1.0000	2.0

```
weights                    0.093536
distance              336312.950000
connection_age_diff      340.000000
steps                     28.000000
dtype: float64
```

Index	weights	distance	connection_age_diff	steps
25%	0.0139	162192.5000	35.0000	11.0
50%	0.0278	231176.5000	92.5000	15.0
75%	0.0586	294623.2500	188.0000	19.0
count	3060.0000	3060.0000	3060.0000	3060.0

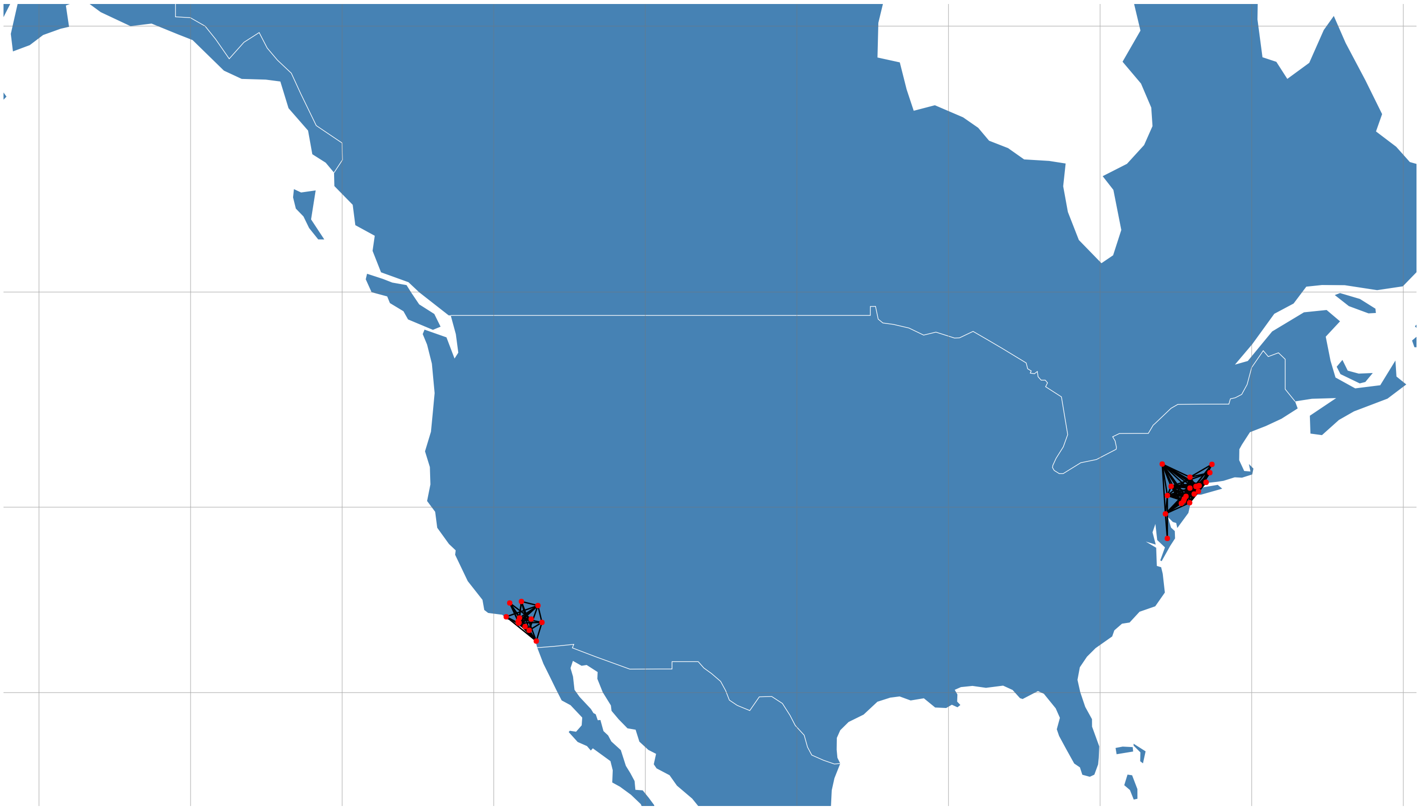
max	0.0982	345838.0000	441.0000	43.00
mean	0.0373	223943.4624	122.3601	15.6

```
No id variables; using all as measure variables
```



Weights

The distribution of connection weights is difficult to describe with multiple peaks, and a stand alone dominate bin weight between 9 and 10%. We can review these highly weighted connections by plotting the subgraph of high weighted connections.



As expected, the peak in connections with weights >0.09 come from the densely connected communites in the highly populated regions of the network, giving further support to the likely affinity of the network to densely populated regions

Distances

The distribution of distances appears to steadily increase towards the connection cutoff of 346kms, suggesting there is a preference to maximizing the connection distance between SCs as opposed to building many close SCs. This is likely due to the SC network still being in an expansion stage with few mature and closely packed SC regions. It will be interesting to analyze this distribution over time to see if a shift in SC spacing strategy can be observed as the netowkr matures.

Connection Age Difference

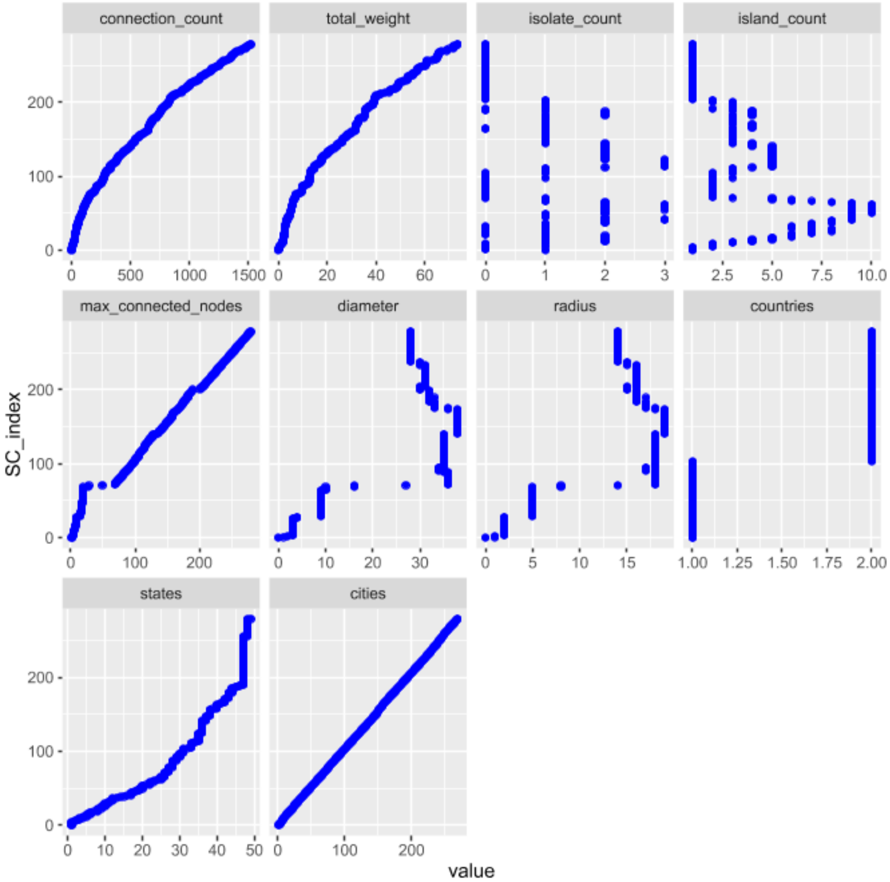
Here "connection age" is defined as the different in SC node index number sorted by date of opening, and is another representation of network maturity. The distribution for the SC network almost linearly decreases as age difference increases, suggesting that most new connections are themselves connected to relatively new connections, where very few new SCs connect to very old SCs. This is again likely due to to the continued expansion of the network outwards as opposed to filling in gaps nearer to older SCs.

Steps

Steps represent the number of direction steps computed by google maps directions API for the directions between 2 SCs. This quanity may be useful in representing the complexity of a connection between 2 SCs. For example, SCs closer to a major highway are likely to have less steps between them, making SC access easier. @5. The distribution of Steps is quite normal, with a slight right tail, which would agree with the assumed preference to lower steps between SCs, but the effect does not appear very to be dramatic.

Dynamic Network Analysis

Index	SC_index	connection_count	total_weight	isolate_count	island_count	max_connected_nodes	diameter	radius	countries	states	ci
25%	69.7500	142.0000	6.6399	0.0000	1.0000	27.7500	16.0000	8.0000	1.0000	26.0000	67.7
50%	139.5000	494.0000	23.3689	1.0000	3.0000	128.5000	31.0000	16.0000	2.0000	36.0000	136.5
75%	209.2500	863.2500	40.2883	1.0000	5.0000	210.2500	35.0000	18.0000	2.0000	47.0000	203.2
count	280.0000	280.0000	280.0000	280.0000	280.0000	280.0000	280.0000	280.0000	280.0000	280.0000	280.0
max	279.0000	1518.0000	73.4814	3.0000	10.0000	280.0000	38.0000	19.0000	2.0000	49.0000	269.0
mean	139.5000	560.1821	26.9274	0.8393	3.5000	131.0786	26.3607	13.4429	1.6286	33.7571	135.1
min	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	0.0000	0.0000	1.0000	1.0000	1.0

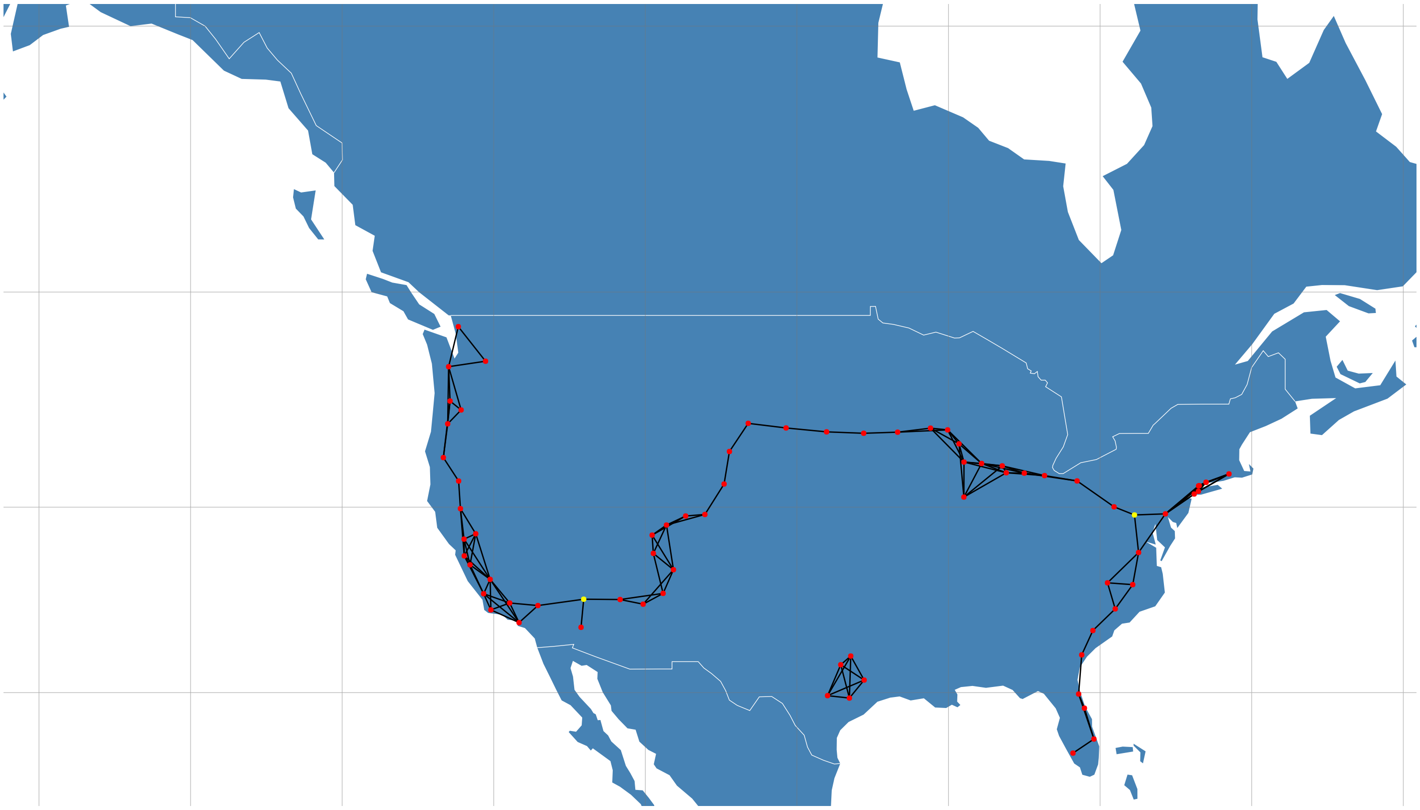


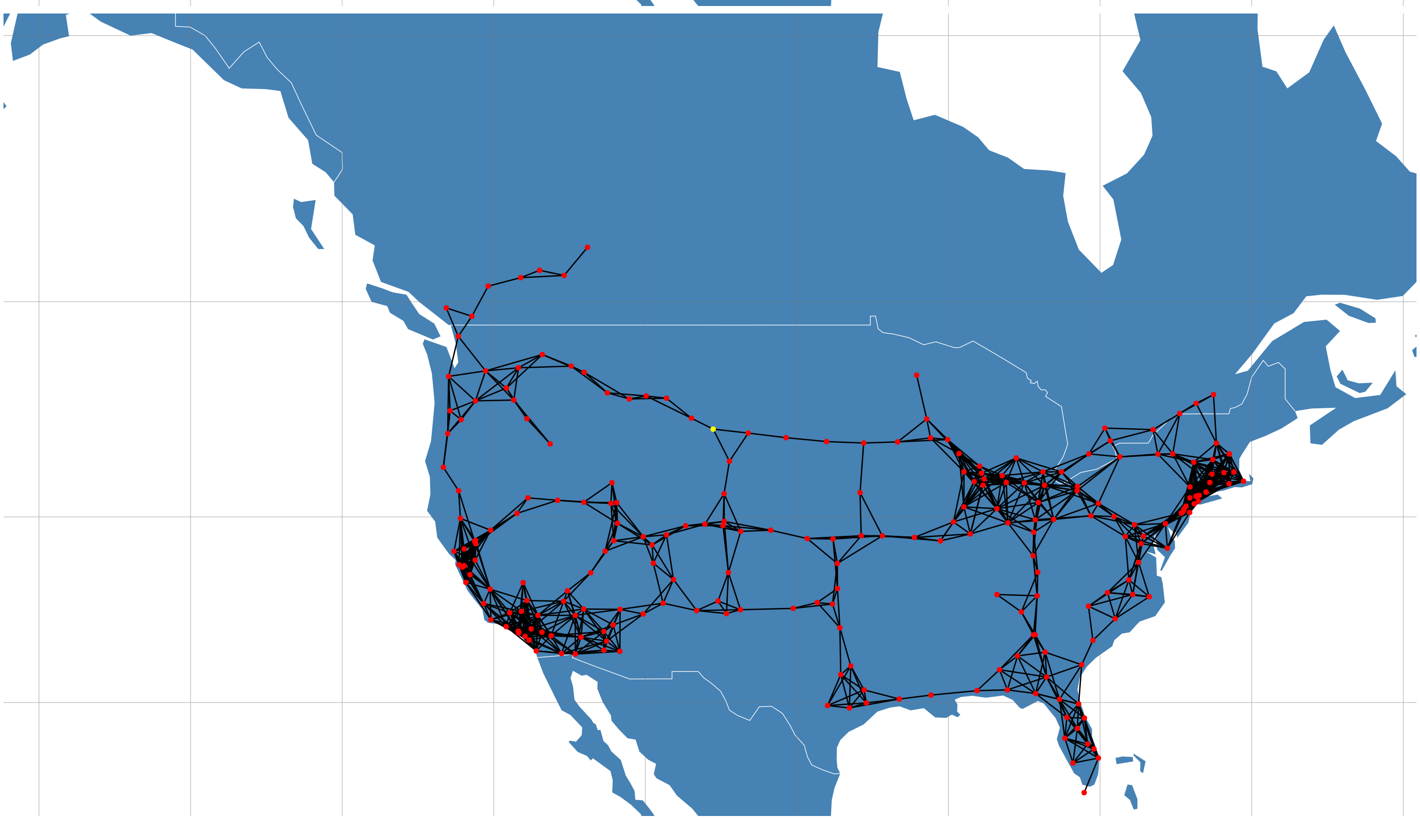
Dynamic Network

Connection counts and the weights between them seem to following a smooth exponential growth trend, which is expected of a complex network. The rest of the plots seem to depict 3 distinct growth periods in the network:

1. Network expansion - Far reaching network expansion with linear growth in the network radius, diameter and disconnected graphs.
2. Network maturation - A flattening and ultimately decreasing diameter/ radius and disconnected graphs.
3. Network equilibrium - All nodes being connected in a single graph with a constant network diameter and radius.

Below we use the network graphic to explore the state of the network at each of the above transitionary periods.

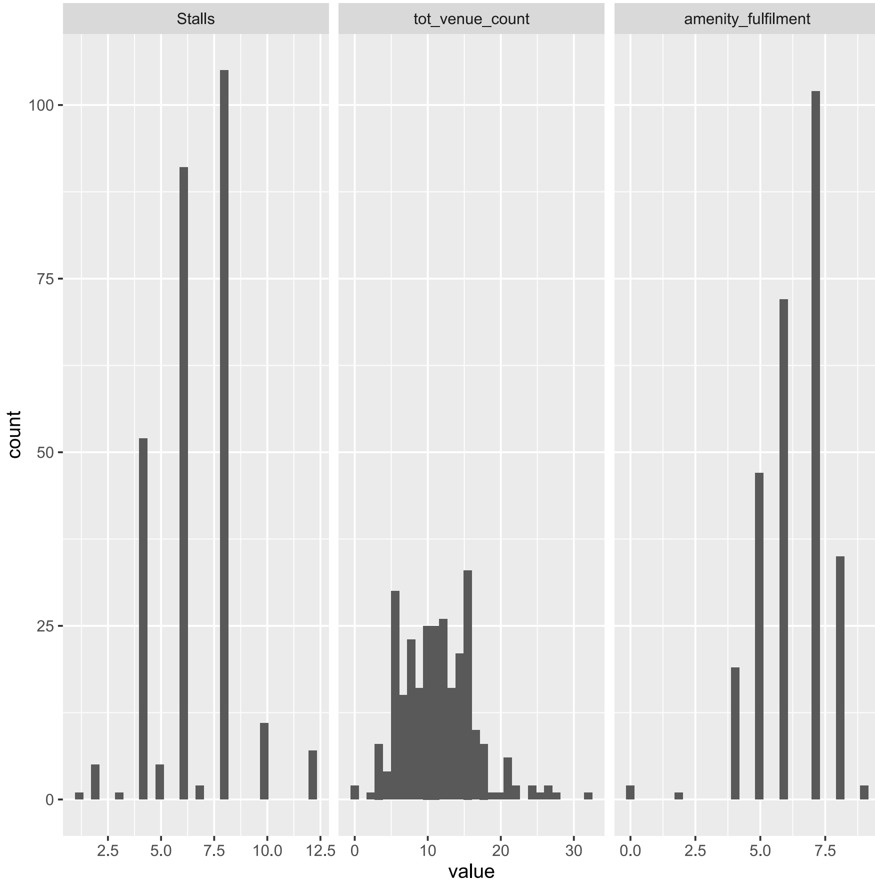


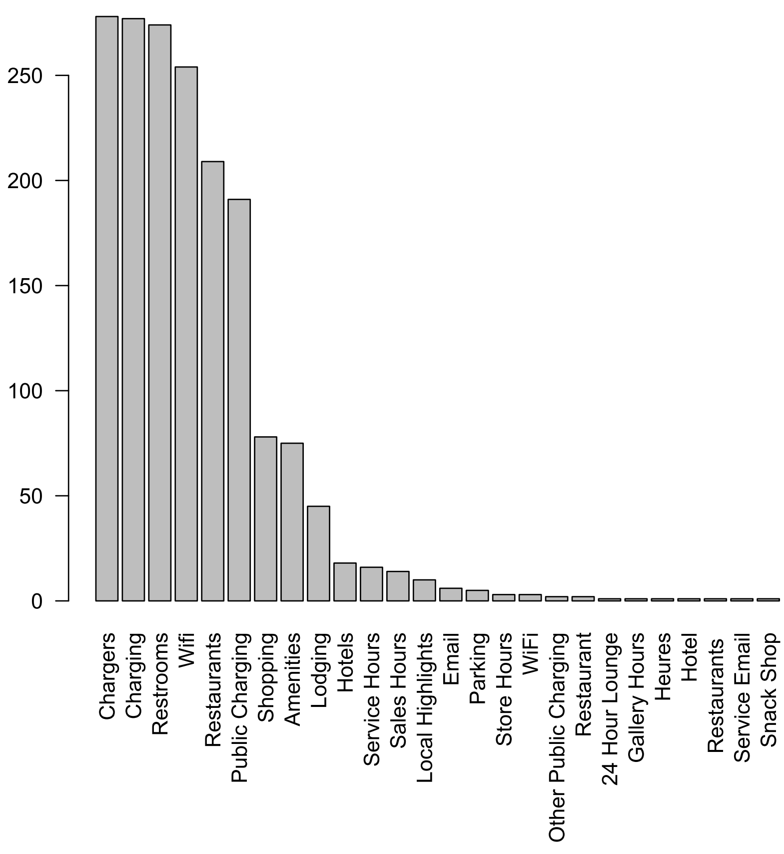


Super Charger numbers 72 and 73 are very important in the growth of the network, as they act to complete the connection of West SCs to East SCs across America. After the connection of these nodes, the network becomes completely dominated by this gigantic sub graph, and the growth of the network becomes focused on the expanison of this gigantic sub as opposed to the expansion to distant and isolated sub structures. The diameter of the main sub graph remains nearly constant, until SC 239 is added which acts to short-circuit the connection of Northwest SCs to the rest of the network. Once this short-circuit is completed, the networks diameter and radius stabalize and it is possible that the current network is now large enough that an equilibrium has been reached between network frontier expansion and filling in short-circuiting pathways in the networks internal structure. Which ever model is utilized to represent the SC network will need to be able to account for these growth transition stages.

- Super Charger Amenity Analysis

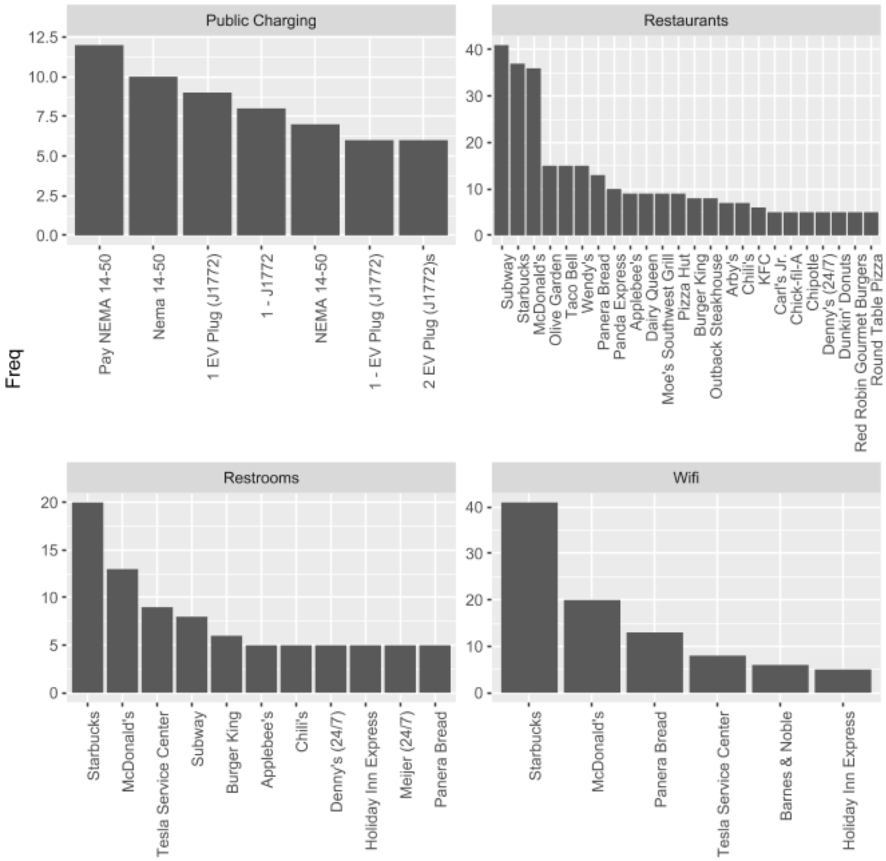
```
No id variables; using all as measure variables
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```





Amenity Types

All Super Chargers have chargers (which is good). The most commonly fufilled amenity type is restrooms, followed by WiFi and then restaurants. It is interesting that the need to be connected to the internet is more commonly satisifed then the need for food.



Venues

Starbucks is the most common venue found at SCs, but does not always qualify as a restaurant, leaving subway as the most common restaurant. Starbucks is the most common restroom and Wifi venue, followed by Mcdonald's. Although these results are not surprising, the distributions may be useful in helping decide good options for future SC locations.

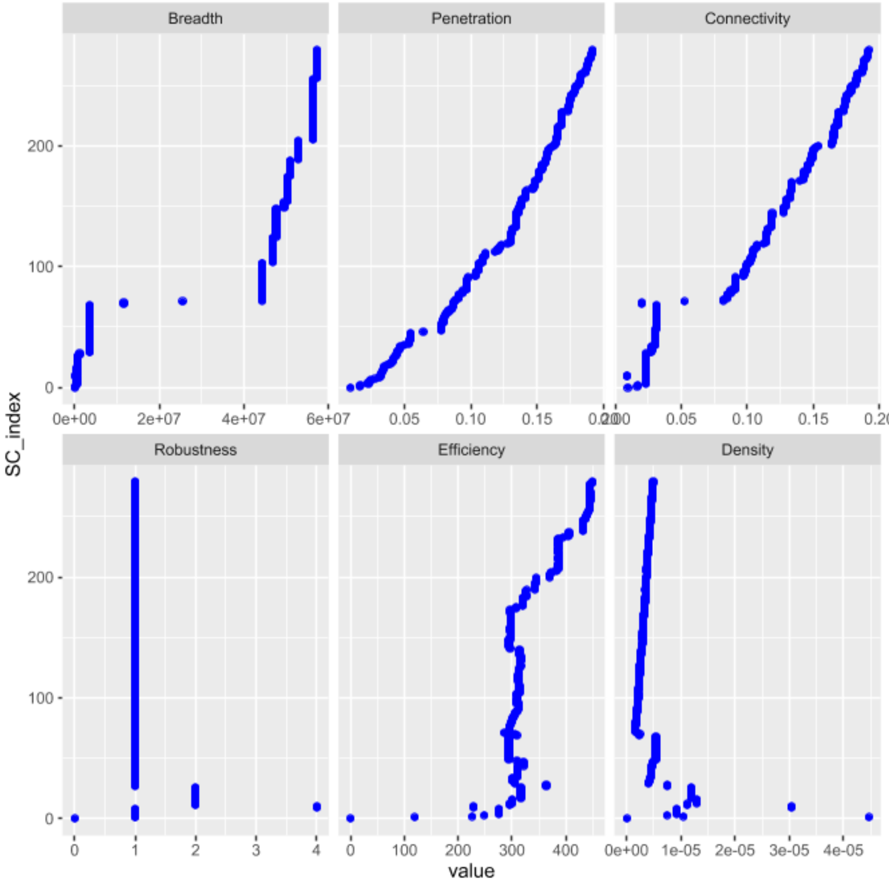
Network Growth Utility

Taking a step back from the detailed network analysis, a higher level motivation for the expansion of the network is considered. Conceptualizing an ideal Super Charger network, one would expect the network to maximize the following characteristics@6:

1. Breadth - the geographic coverage area of the network defined as the area of the ellipse formed by the haversine distances of the network's furthest North and South and East and West connected Super Chargers.
2. Penetration - the % of the population represented in the network defined as: (net_pop/tot_pop)
3. Connectivity - the % of population represented in the main connected sub graph of the network
4. Robustness - the number of SCs that can be offline before the network becomes disconnected defined with networkx's node_connectivity algorithm
5. Efficiency - the avg effective travel distance of a trip using a SC connection, defined by the sqrt(Breadth/pi) divided by networkx's average_shortest_path_length algorithm
6. Density - the average number of SCs per km^2 in the network defined as Super Charger count of network Breadth

These characteristics are analysed through the evolution of the current Super Charger network below.

Index	SC_index	Breadth	Penetration	Connectivity	Robustness	Efficiency	Density
25%	69.7500	11608471.6282	0.0866	0.0308	1.0000	297.7572	0.0(
50%	139.5000	47612601.0673	0.1341	0.1183	1.0000	313.5384	0.0(
75%	209.2500	56367655.5581	0.1653	0.1653	1.0000	384.3292	0.0(
count	280.0000	280.0000	280.0000	280.0000	280.0000	280.0000	280.0(
max	279.0000	57173044.4252	0.1918	0.1918	4.0000	449.8908	0.0(
mean	139.5000	38759257.0991	0.1232	0.1114	1.0750	335.6675	0.0(



Growth Utility Characteristics

These plots provide insight into the characteristic contributions to the network the addition of a SC provides. Most measures appear to be being maximized, but at much different rates per SC addition. Penetration and Connectivity appear to have linear or sub linear growth that may further decelerate forming a logarithmic trend, as the high population density locations are selected and each additional Super Charger adds continually less new population representation to the network. Breadth also appears to have a linear growth profile in the network maturation stage of the network but appears to be slowing down through the current equilibrium stage of the network. It's apparent that increasing the Breadth is a desirable characteristic of adding a new SC, but does not appear to be the dominant deciding factor on a SC addition. Density also presents a clean linear trend through the maturation and equilibrium growth stage of the network, without any indication of decelerating. It is intuative that as the network expands, and population representation approaches 100%, a larger focus on network Density will be used to determine the addition of a future Super Charger. This is also apparent in the more sporadic, but still increasing Efficiency plot, which was fairly flat through the network maturation growth stage, but has acceleted through the equilibrium growth stage of the network. The only characteristic that does not appear to be maximized is the networks robustness, which indicates that, on average, the removal of a single SC in the network can cause SCs to become disconnected from the rest of the network. Again, this is somewhat intuitive, as although having a robust network is desirable, this is a characteristic of very mature networks where building in redundancy is more important that gaining fractionally more network coverage. This information should be useful in building a network growth model that will need to accurately represent these growth characteristics.

Conclusions

EDA Summary

The first conclusion obtain from this analysis is that complex networks are...complex; with nearly an infinite amount of ways to slice and dice their representation, characteristics and dynamics. But taking a deep dive into a networks dataset provides the development of an intuition for a network, which will provide robust "gut checks" about interpretations, predictions and inferences of uncertain future network growth. Beyond the "meta" learning from the analysis, a few key take aways about the network were discovered:

1. The Super Charger Network has a log-normal degree distribution which is not well represented by many well established complex network generation models
2. The network is not very robust, with a single SC connecting 30% of the network population through their shortest paths
3. Super Chargers are attracted to high population density locations, and SC communites form around these population centers
4. Population density is not the only part of the story in determining the growth of the network, with many states being drastically over and under represented by SCs when compared to their populations
5. The network appears to have undergone 3 distinct transitionary growth stages: expansion, maturation and equilibriation
6. The typical Super Charger will have 7 nearby venues that suppy restrooms, Wifi and restaurants which usually consist of Starbucks, Subway or McDonald's.
7. The rate of network Penetration and Connectivity appears to be slowing down, while Efficiency and Densification continue to rise linearly. Network Breadth and robustness appear to be less important in the current growth of the SC network.

Next Steps

Now that an intuition of the network has been developed, an expansion model needs to be created to build a predictive decision support system in helping decide the optimal locations for future Super Chargers. This model will need to account for the various characteristics discovered above. After a brief literature review, the proposed steps for building this model are as follows:

1. Attempt to directly model the network, possibly with the Multiplicative Attribute Graph Model (https://cs.stanford.edu/people/jure/pubs/mag-im12.pdf)
2. Develop a Multi-Criteria decision support system accounting for network, geographic and SC characteristics to model the utility function of a future node
3. Develop hueristics to narrow the search space of potential future SC locations
4. Implement an optimization for the selection of the optimum next SC location (or next set of SC locations)
5. Validate the model with the held-out dataset of SCs currently under construction
6. Review, document and share

Tool Review

Beaker

This analysis has been an a way for me to experiment with Beaker Notebooks, and I very much enjoyed the experience. The ability to use this notebook as my canvas to throw everything I have in my toolbox to this problem on a single, seamless whiteboard has been fantastic, and I was able to go further and dig deeper by being able to directly combine what I know how to do in Python, R and D3.js. I think the best part is being able to pass data directly from python to D3 and review the visualization immediately. The ability to use python to munge data in any which way and instantly visualize this, inline, with an advanced D3.js graphic allows for almost instantaneous iteration of complex concepts. I will definitely be coming back to Beaker Notebooks for many future projects.

Networkx

Networkx was a very helpful tool for this paticular analysis, and lead me to really learn some of the fundamental concepts of complex networks through. Complex Networks seems to be a rapidly evolving discipline, and even in my brief research review, I found many algorithms not implemented within networkx, such as the MAGfit algortithm for fitting a real network to the MAG model. Perhaps in the future this will provide a perfect opportunity and platform for me to contribute to a substantial open source library.

Holy list-comprehension, Batman!

I'm not sure if it's a result of how I structure the data model, if it's inherent in complex network EDA, or if I just got into a list-comprehesion only mindset, but I sure practiced my list-comp-fu in this analysis. It definitely wasn't the most performant or readable code, but it did get the job done. I think this is a strong indication that I should sharpen my data-munging skills; learn some higher-level techniques and explore the capabilities of pandas and R packages. I have completed Udacity's Data Munging (https://www.udacity.com/course/viewer#!/c-ud651/l-729069797/m-804129329) course, and it might be time for a refresher.

But first, I have a Super Charger network to build…