**Name: Amir Kazi  
Policy Area: Transportation  
Concept: Change in Demand for Public Transportation & Taxis by Variation in Chicago Weather**

**Idea:**

Taking public transportation (CTA buses and trains) in Chicago often requires walking to a bus stop/train station from location A, or walking from one to location B.   
Taxis, on the other hand, tend to drop off at a precise location, with less need for on-foot travel.

It is also conceivable that it is far more inconvenient to walk to or from bus stops/train stations during days of inclement weather (rain/snow/hail) compared to days of good weather (sunny, cloudy etc.).

I wish to research (and visualize) if there are changing patterns in demand for public transportation and taxis with these variations in weather.

Possible questions/visualizations[[1]](#footnote-1):

1. CTA ridership by day.
2. Taxi ridership (by day/hour/15-minute block). Additionally, demographics of demand by area (due to higher costs associated with cabs) for example, more taxis going north than south Chicago.
3. Taxis taken to or from bus stops / train stations (as an intermediary in a transit)
4. Where are taxis going and from where?
5. Changes to all the above based on weather

**Data Sources[[2]](#footnote-2):**

1. Taxi Trips
   1. Taken from Chicago Data Portal (<https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew/data>)
   2. Updated every 15 minutes with new rides.
   3. Contains timestamps for trip start and end, geolocation (latitude and longitude) of pickup and drop-off), census & community tract, fare, and taxi id.
   4. 113 million rows, 23 columns.
2. CTA ridership: Bus Routes – daily total by route
   1. Taken from Chicago Data Portal (<https://data.cityofchicago.org/Transportation/CTA-Ridership-Bus-Routes-Daily-Totals-by-Route/jyb9-n7fm>)
   2. Has daily data as opposed to hourly data.
   3. Variables: bus route, date[[3]](#footnote-3), number of total rides in a day.
   4. This doesn’t identify where the rider got on board the bus.
   5. 800 thousand rows, 4 columns.
3. CTA ridership: L Station Entries- daily total
   1. Taken from Chicago Data Portal (<https://data.cityofchicago.org/Transportation/CTA-Ridership-L-Station-Entries-Daily-Totals/5neh-572f>)
   2. Has daily data as opposed to hourly data.
   3. Variables: station id, station name, date, number of total rides in a day.
   4. This doesn’t identify where the rider went (i.e. on which train) after they entered the station.
   5. 923 thousand rows, 5 columns.
4. Weather: Chicago Weather (Historic and up to date)
   1. As of right now, weather data is unavailable on US Government websites due to ‘lack of appropriations’. From Google Datasets, I found a Kaggle dataset that has Chicago weather hourly data from 2012 – 2017. https://www.kaggle.com/selfishgene/historical-hourly-weather-data#weather\_description.csv
   2. The above dataset contains, amongst other things, ‘weather descriptions’ by the hour which can give weather information such as ‘light rain’ or ‘overcast clouds’.
5. Shapefiles for Chicago (by neighborhood, ward, census tract, zip code)
   1. <https://data.cityofchicago.org/browse?tags=shapefiles>
6. Shapefiles for CTA L Rail lines
   1. https://data.cityofchicago.org/Transportation/CTA-L-Rail-Lines-Shapefile/53r7-y88m
7. Shapefiles for CTA Bus
   1. Bus routes: <https://data.cityofchicago.org/Transportation/CTA-Bus-Routes-Shapefile/d5bx-dr8z>
   2. Bus stops: <https://data.cityofchicago.org/Transportation/CTA-Bus-Stops-Shapefile/pxug-u72f>

**Theory & Research Papers:**

1. Guo, 2007. The impact of weather on transit ridership in Chicago. <https://docplayer.net/620895-The-impact-of-weather-on-transit-ridership-in-chicago.html>
2. Nesse, 2012. Impacts of Inclement Weather on Transit Ridership: A Minneapolis, MN Case Study. <http://www.transportchicago.org/uploads/5/7/2/0/5720074/ps3_transitinclementweather.pdf>

**Questions/Notes for Alex:**

1. Taxis data is within 15-minute blocks, but CTA data is for a day. For weather conditions like rain (which may not last all day), it is harder to track demand variation for CTA.
2. While taxi data has pickup and drop-off locations, L-train data only has pickup location, while bus data has no pickup nor drop-off (just the general route).
3. Given the two issues above, what should I do about Trains and Bus data? Bus data seems least useful, should I drop it? What about L (train) data? Should I just compare Taxi data (with itself as a control, i.e. only accounting for weather changes)?

1. More may be added. [↑](#footnote-ref-1)
2. Expansive data source: <https://www.transitchicago.com/data/>

   Chicago Data Portal: <https://data.cityofchicago.org/browse?category=Transportation> [↑](#footnote-ref-2)
3. Can use Lubridate to get other information from date. [↑](#footnote-ref-3)