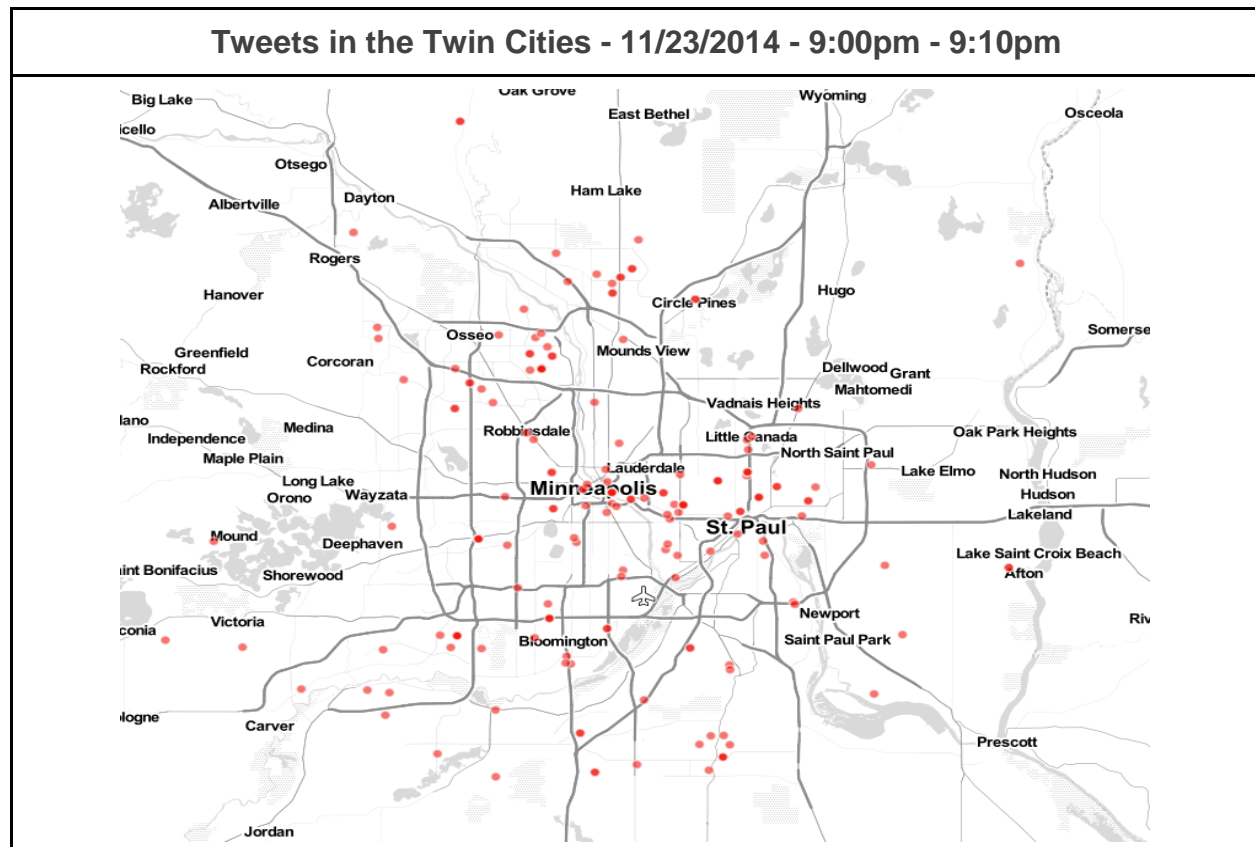


# Lab 4

## Part 1

### Analyze tweets from the Twin Cities

1. How many geotagged tweets are coming in per minute? (Collect tweets for at least 10 minutes). Also note what time and day you collected this info.
  - a. **26.5 geotagged tweets per minute in Minneapolis on Sunday night (Nov 23) from 9:00pm to 9:10pm.**
2. What is the most common hashtag (word preceded by #) in your dataset?
  - a. **We collected this data during the American Music Awards, which took place on November 24th from 7 to 10pm. The most common hashtag was #AMAs**
3. Ask some other question of your Twin Cities dataset (that is relatively interesting and easy to answer). Report your question and results.
  - a. **We wanted to see the spatial distribution of the tweets from Minneapolis during this time period. Because some of the tweets weren't geotagged but had a location associated with them, tweets from outside the bounding box area we also grabbed. This is because Twitter uses polygons (name, type) to describe place names. If some of these polygons overlapped our bounding box, the tweets from that place we also included in the result. Below is a figure showing the spatial distribution of the tweets.**



4. What are the key lines of code you used to collect your tweets? Copy and paste these lines into your assignment document

#### Key lines of code (using Tweepy - twitter api library for python)

```
# Stream
class StdOutListener(tweepy.StreamListener):
    ''' Handles data received from the stream. '''

    def on_status(self, status):
        # grabs different tid-bits from tweet and stores them in various variables
        try:
            tweet = status.text.encode('utf-8')
            tweet = tweet.replace('\n', '\\n')
            user = status.author.screen_name.encode('utf-8')
            userid = status.author.id
            tweetid = status.id
            time = status.created_at
            source = status.source
            geoEnab = status.user.geo_enabled
            # grabs coordinates if the exist
            try:
                Coords.update(status.coordinates)
                LongLat = (Coords.get('coordinates'))
            except:
                LongLat = [0,0]
            # grabs place info if it exists
            try:
                placeName = status.place.full_name
                placeType = status.place.place_type
            except:
                placeName = "none"
                placeType = "none"
            # grabs user location if it exists
            try:
                userLoc = status.user.location
            except:
                userLoc = "none"

            # prints stored info to a tab-delimited text file
            print >>file, userid, "\t", user, "\t", time, "\t", tweetid, "\t", tweet, "\t", source,
            "\t", LongLat[0], "\t", LongLat[1], "\t", userLoc, "\t", geoEnab, "\t", placeName, "\t", placeType

            # prints username to console to show progress
            print user

        except Exception, e:
            print >> sys.stderr, 'Encountered Exception:', e
            pass

#####
# further down the script...
#####
# Main section of script
if __name__ == '__main__':
    listener = StdOutListener()
    auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_token, access_token_secret)
```

```
# initiates the twitter stream
stream = tweepy.Stream(auth, listener)

# filters tweets in stream by hashtag
stream.filter(locations=[-93.65,44.75,-92.90,45.13])
```

## Analyze tweets by keyword

1. What percent of tweets for each keyword are geotagged?
2. What percent of tweets for each keyword have a non-null user location field?

Answers to 1 & 2 (after 25 minutes of collecting at 11pm on Sunday night):

	Total Tweets	% Tweets Geotagged	% Tweets w/ non-null User Location
#Vikings	20	0%	55.0%
#Packers	115	2.6%	47.8%
#YOLO	130	3.1%	44.6%

3. Do you see any non-geographic information in these location fields? If so, provide some examples.

### Examples of non-geographic information in user location fields:

- @misshudgens in instagram
- #Squaddd
- |-/
- \$\$\$
- 16-03-14
- 337 2 504: Belly Of The Beast
- And Extra Points 4 Everything@
- asdfghjklove
- Conquering Kingdoms
- Finding that Happy Place
- Google Play Store - Free App
- Graphic Designer/Pubstomper
- Internet (everywhere!)
- PROXIMO PRESIDENTE DE MEXICO
- World Wide Web

## Look a little closer...

1. Do you see any other possible sources of geographic information? If so, what? Consult the Twitter API documentation and explain in a few sentences or two the nature of this information. What you believe to be its positives and negatives relative to geotags?
  - a. **Places can be another source of geographic information - aside from geotagging a tweet. Places are objects that can be grabbed through the Twitter Streaming API. These are specific name locations with a corresponding bounding box or coordinates. Tweets that are associated with a particular place object are not**

necessarily being tweeted from that location, but could also be about that particular place. Places are good because they offer more location information on tweets that would otherwise not be spatial at all. Personally, we also think that Twitter users are more inclined (they feel safer) adding a place location rather than using their exact location to geotag the tweet. Some negatives are that people can type in a place name and not actually be anywhere near that place.

- b. Other sources of geographic information can be content based - this is a concept we discussed in class where the content or hashtags of the actual tweet can be used to determine a user location. For example, we could take all the #vikings tweets and assign them to Minnesota and take all the #packers tweets and assigned them to Wisconsin. This of course can cause all sorts of problems because football fans are not limited to state boundaries, but content based geolocation can work in certain other ways.

## Part 2

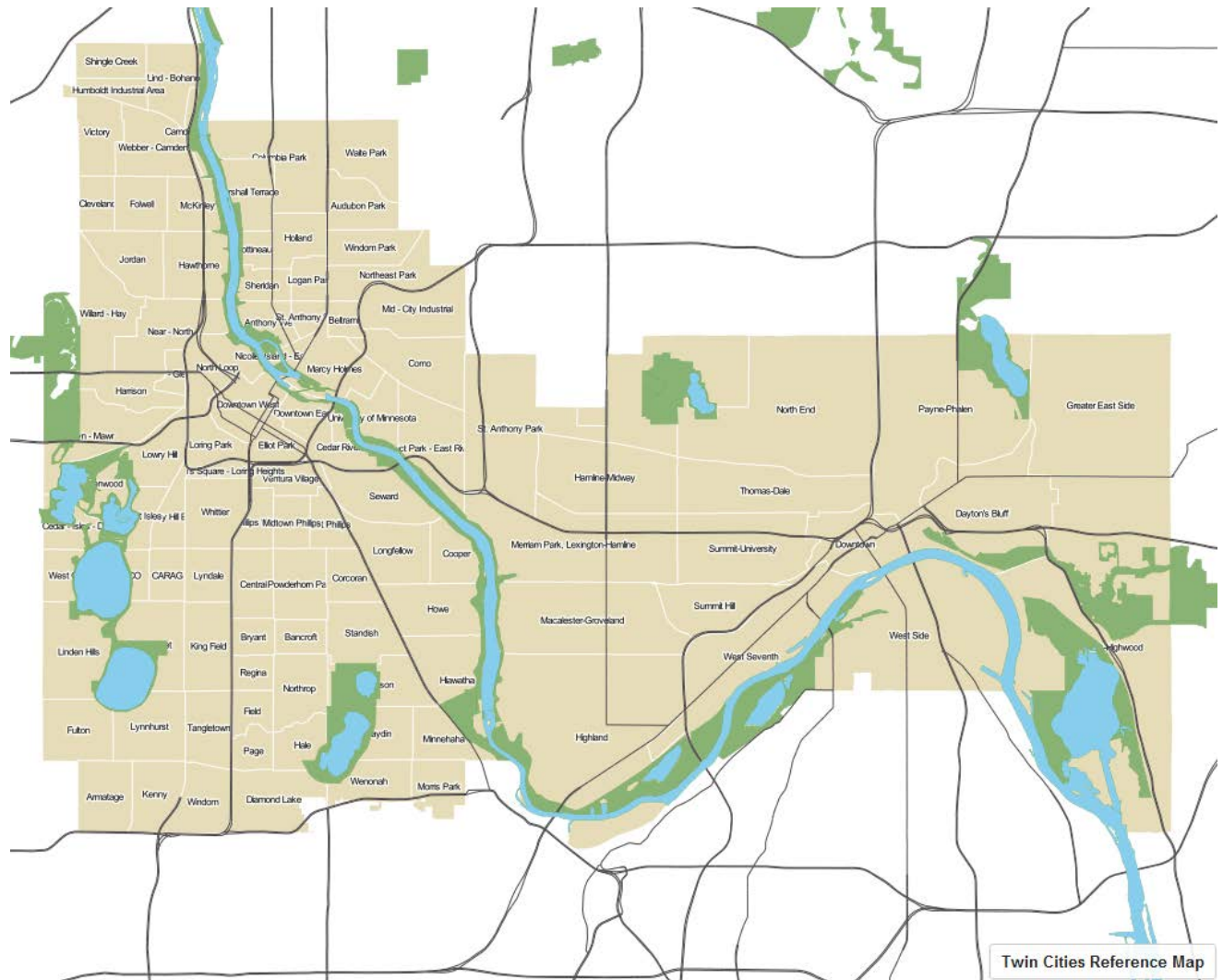
### Task #1: Make a super ugly map



## Task #2: Make a basic reference map of the Twin Cities

Click on the image below or link to redirect to the interactive map hosted on MapBox.

<http://a.tiles.mapbox.com/v3/agatamiszczyk.TwinCitiesReferenceMap/page.html#12/44.9701/-93.2087>



### Task #3: Make a basic thematic map of the Twin Cities

We mapped the number of people who live alone. For this particular map, we went with the Natural Break classification scheme. The data is showing raw numbers (not standardized for population) so this classification will pick out areas of high population. because there is a greater chance that someone will live alone in a high density area.



**Task #4: Publish this map to your MapBox account and include a link in your homework document**

<http://a.tiles.mapbox.com/v3/agatamiszczyk.TwinCitiesThematicMap/page.html#10/44.9308/-93.6626>

