# Pick a projection in order to transform the data into 2d M

# Measure the Euclidean distance in the projected space.

1. In addition to your presentation, I expect you to submit the ipython notebook you created while preparing your presentation. I need to see ALL the code you wrote (not necessarily in the presentation)

2. Your presentations need not be extremely text-heavy, because the full details of your research.

You can think that you need to prepare 6-10 powerpoint

slides (depending how much info you put per slide), each slide giving information about a different part of your analysis. The number of slides does not matter, but your presentation of the material does.

3. For Wed, April 29, you should prepare a 5 min (really! 5 minutes only!) explanation of the most interesting things you found

in the course of your research.

One of the instructors (me, Harry) will approach you, listen to your explanation, and possibly also ask you a few questions.

This discussion with your instructors will be part of your final grade.

4. Your slides should contain information about the dataset you collected (and how you collected it).

Some basic statistic's of your dataset. Then you should describe your hypotheses

and what you did in order to be conviced that your hypotheses holds or does not hold.

5. On Wed 4pm in the Hariri Institute, MCS180, to set up the your computers. I expect we will conclude the poster session at 6PM.

6. During the session, you should alternately stand in front of your laptop and be ready to explain it to people that come by,

and visit other people's laptops, as in any scientific poster session.

I am pretty certain that some of things that your classmates discover are not known to the general public, so this is a

great way to get some knowledge that you can use to impress future employers and colleagues.

Hypothesis:

Is there a trend in x in regards to y, and if so, what is that trend?"

What factors attribute to higher occurrences and/or spread of the West Nile virus?

Is there a trend in the spread and/or frequency of the west Nile virus in regards to mosquito, species, weather, and insect spray?

At what rate did the virus spread? Which factors affected the spread and frequency?

(Chicago)

Results/Analysis:

1) Basic description with frequency graphs and statistic of the what the data entails

2) Clustering

-use k-means++

-easy to implement

-provable approx guarantee just from the first iteration

(initialization). More iteration improves cost

-works well in practice

-

-In this context, we're clustering and k-means++ using the euclidean algorithm

to group data objects in respective group in which objects are simiar to each other

Spatial cluster analysis plays an important role in quantifying geographic variation patterns.

-interpolate between two methods.

-D(x) be the distance between x and the nearest center selected so far

-choose next center with probability proportional to (D(x))^a = D^a(x)

-random initialization, furthest-first traversal, k-means++

3) Elbow graph to determine the number of clusters

4) Heat map. Great visualization to for spatial analysis plays an important role in quantifying geographic variation patterns.

5) Graph percentage of traps held mosquitoes that were infected.

Trap infection rate for 2007, 2009, 2011, 2013 using the train dataset

Highest percentage in 2007 and 2013 compared to 2009 and 2009

6) Graph which mosquito species may have had the highest rate of carrying the West Nile Virus

7) Graph of how weather has affected the frequency and spread of the West Nile Virus

8) Graph of how spray has affected the frequency and spread of the West Nile Virus

(compared to the Heat Map by layering graphs)

9) Logistic regression?

10)