

Weather influence on Human Mobility

A Twitter study

Henrique Dantas Emanuel Dias Liang Huo Carryoles Maria
Vadim Weis

TU Delft

February 13, 2013

Table of contents

1 Introduction

- Related research and state of art
 - Hypothesis

2 Methods and Materials

- Twitter
 - Weather
 - Collecting and Storing Data
 - Collecting Twitter Data
 - Collecting Weather Data
 - Data Analysis

3 Results

4 Discussion

Introduction

- How to identify different factors that influence our daily routine
 - Study the relation between human mobility and weather conditions
 - In particular, analyze the weather on popular routes

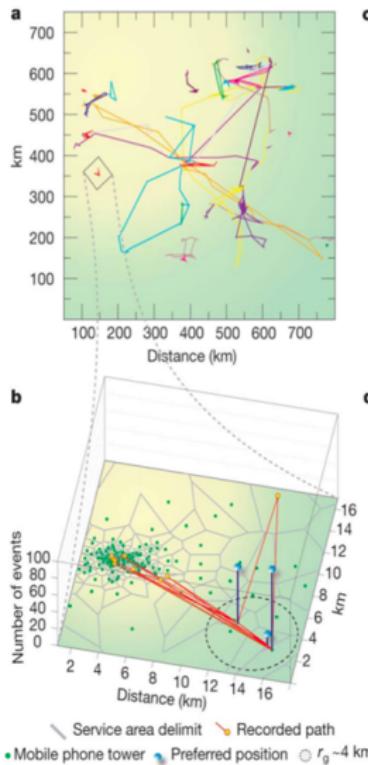
Related work

Academic research on human mobility is based on personal data collected from:

- Cellular Networks
 - Social Networks
 - Or a combination of both

Personal data from cellular network

- Understanding human mobility
 - Authors disregard influence of external factors and weather conditions
 - Personal data is acquired by saving the location as given by the base station



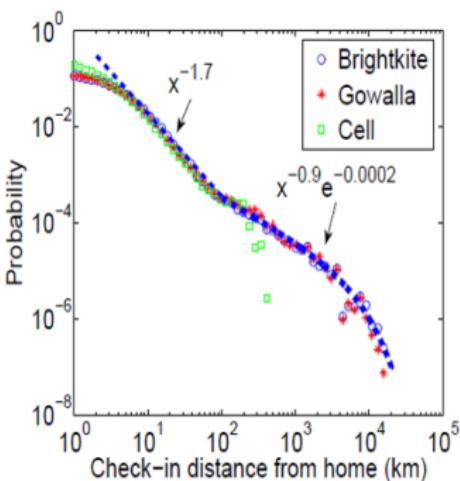
Personal data from Social Network

- Finding your friends and following them to where they are
 - Authors assume friends have a major influence on human mobility
 - FLAP (Friendship Location and Analysis Prediction)



Personal data from Social Network

- Friendship and Mobility: User Movement In Location-Based Social Networks
 - Authors assume friends have a major influence on the traveling distance
 - Similar patterns emerge regardless of data set



Hypothesis

- Based on individual experience, not only the weather impacts the decision to travel
 - Going to school and to work is mostly mandatory
 - Primary hypothesis: *The weather conditions have a dominant effect on human mobility*

- What is twitter?
 - Why people use twitter?

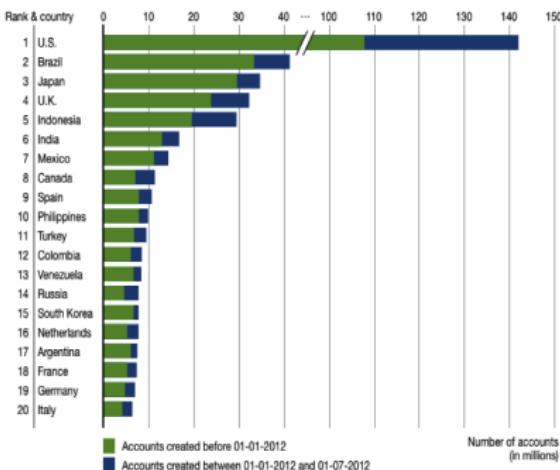


Twitter accounts

- 16.8 million Dutch inhabitants (28 Dec 2012 CBS)
- approximately 50% of Dutch inhabitants have a twitter account

Top 20 countries in terms of Twitter accounts

(accounts created before 01-07-2012)

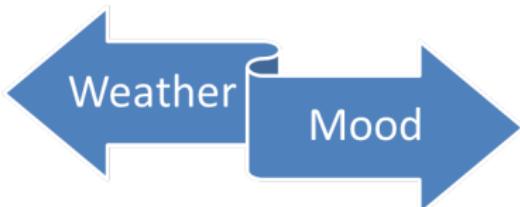


Twitter

- Advantages
 - Built on open source software
 - REST and Streaming API
 - REST is limited
 - Streaming
- Data is in JSON format
- GSON library used for conversion
- Stored in text file and processed with Apache Pig

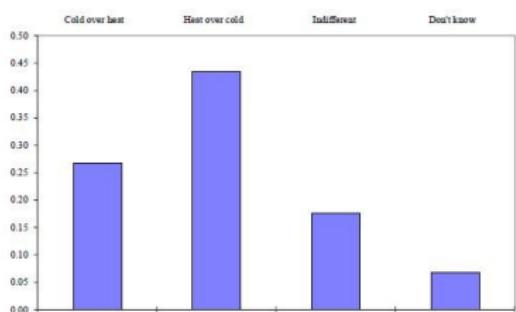
Weather & Mood

- Hours of Sunshine
 - Precipitation
 - Temperature
 - Wind direction
 - Wind Velocity
 - Humidity
 - Barometric Pressure

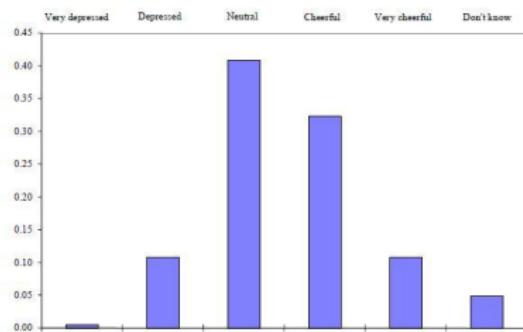


- Anxiety
 - Skepticism
 - Optimism
 - Concentration
 -

General Preferences of Dutch to Weather



(a) Weather preference in Summer



(b) Attitude towards a mild winter

Figure from Richard S.J. Tol (ed.), Weather impacts on natural, social and economic systems in The Netherlands

Mobility prediction via Weather information

Weather is correlated with many aspects of our life

Twitter data with geo tag & Weather among most popular routes

Most influential weather factors, Mobility Prediction

Measurement

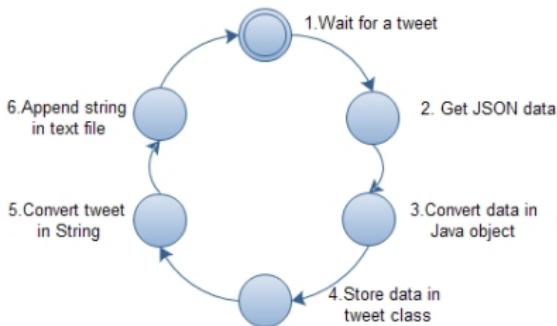
- Measurement Period
 - 6 Dec - 8 Jan
 - Total tweets, including noise, 3769222 ~ 4 million
 - Highest amount (1st January) - 142160
 - Average Day - 112581.9 tweets

Collecting and Storing data

- Try to collect only tweets from NL
- Store format

	User		Tweet		Geo		Tweet	
Count	UserId	Screen_name	msg_id	Created_at	Lat	Lon	Source	Text

- Conversion Process



Select monitoring weather stations |



Select monitoring weather stations II



Weather data from KNMI I

Climatology

Daily weather data of the Netherlands

Choose station, year, month, day and then push the button "show"

Weather data of friday 18 January 2013 at De Bilt

Temperature

	Average	Precipitation	
Mean	-1.7 °C	3.3 °C	24h sum 0.0 mm
Maximum	-0.1 °C	5.9 °C	Duration 0.0 hours
Minimum	-3.4 °C	0.6 °C	

Sun, cloud cover & visibility

Duration sunshine	1.8 hours	
Relative sunshine duration	22 %	25 %
Average cloud cover	7 octa's	
	cloudy	
Minimum visibility	4.3 km	

Wind

Mean	4.0 m/s	= 3 Bft
Maximum hourly mean	6.0 m/s	= 4 Bft
Maximum gust	10.0 m/s	
Prevailing direction	94 °	= E

Relative atmospheric humidity

Mean	80 %	87 %	Mean air pressure	1009.5 hPa
------	------	------	-------------------	------------

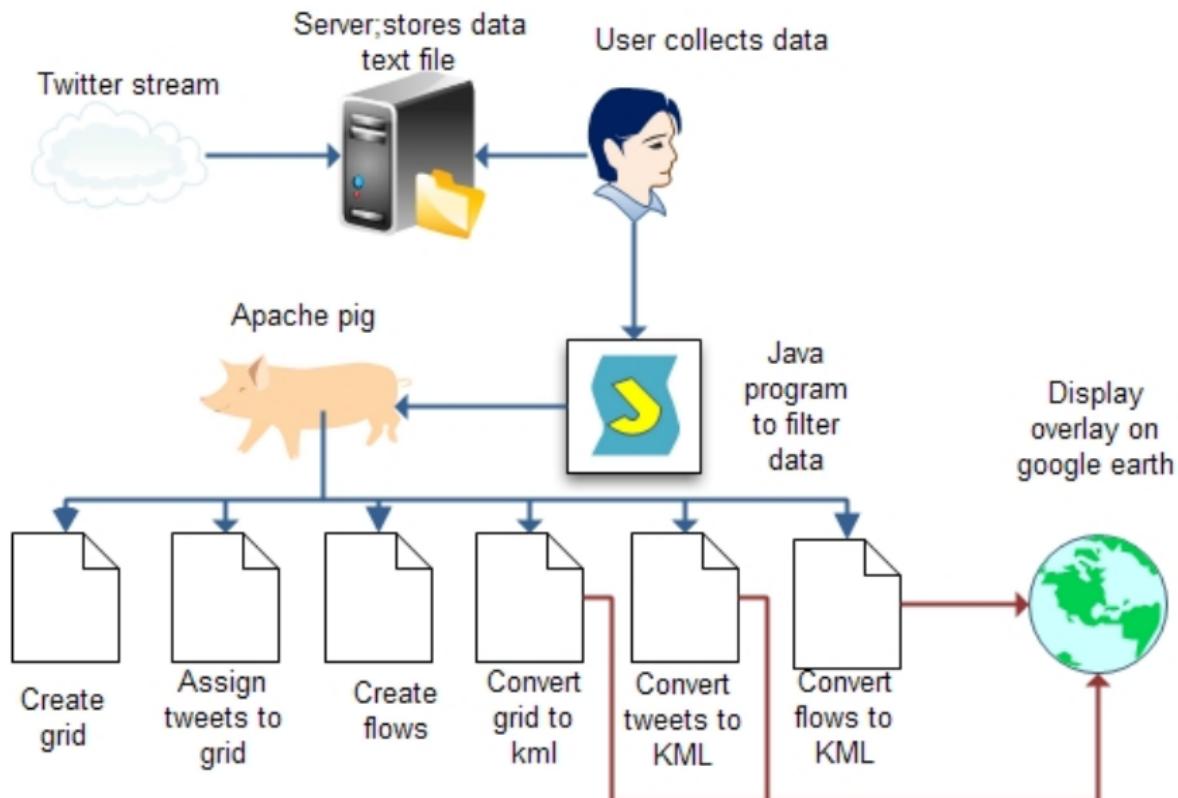
To see more elements of these stations, select "Download"

Explanation Download Nederlands

Weather data from KNMI II

```
# SOURCE:  
# ROYAL NETHERLANDS METEOROLOGICAL INSTITUTE (KNMI)  
#  
# STN      LON(east)    LAT(north)     ALT(m)   NAME  
# 375:      5.706        51.657       21.10   VOLKEL  
#  
# YYYYMMDD = Date (YYYY=year MM=month DD=day);  
# FG      = Daily mean windspeed (in 0.1 m/s);  
# TG      = Daily mean temperature in (0.1 degrees Celsius);  
# SQ      = Sunshine duration (in 0.1 hour) calculated from global radiation (-1 for <0.05 hour);  
# RH      = Daily precipitation amount (in 0.1 mm) (-1 for <0.05 mm);  
#  
# STN,YYYYMMDD,    FG,    TG,    SQ,    RH  
#  
375,20121217,    35,    64,    5,    31  
375,20121218,    25,    52,    2,    -1  
375,20121219,    21,    47,    0,    0  
375,20121220,    56,    22,    10,   73  
375,20121221,    22,    53,    0,    21  
375,20121222,    42,    72,    0,    155  
375,20121223,    68,    113,   0,    215  
375,20121224,    52,    101,   0,    43  
375,20121225,    68,    92,    12,   80  
375,20121226,    68,    76,    12,   39  
375,20121227,    47,    76,    27,   29  
375,20121228,    43,    64,    3,    25  
375,20121229,    64,    106,   32,   14  
375,20121230,    93,    75,    18,   23  
375,20121231,    92,    91,    0,    20  
375,20130101,    51,    63,    5,    134  
375,20130102,    46,    54,    57,    7  
375,20130103,    50,    95,    0,    15  
375,20130104,    42,    92,    0,    9
```

Data Analysis



Pre-processing

Exclude tweets:

- ① With no geo tags
- ② From users with < 2 tweets
- ③ Published on weekends
- ④ Located outside NL

Then remove undesired whitespace from message.

Total : ~1.5 M tweets

Quad-tree

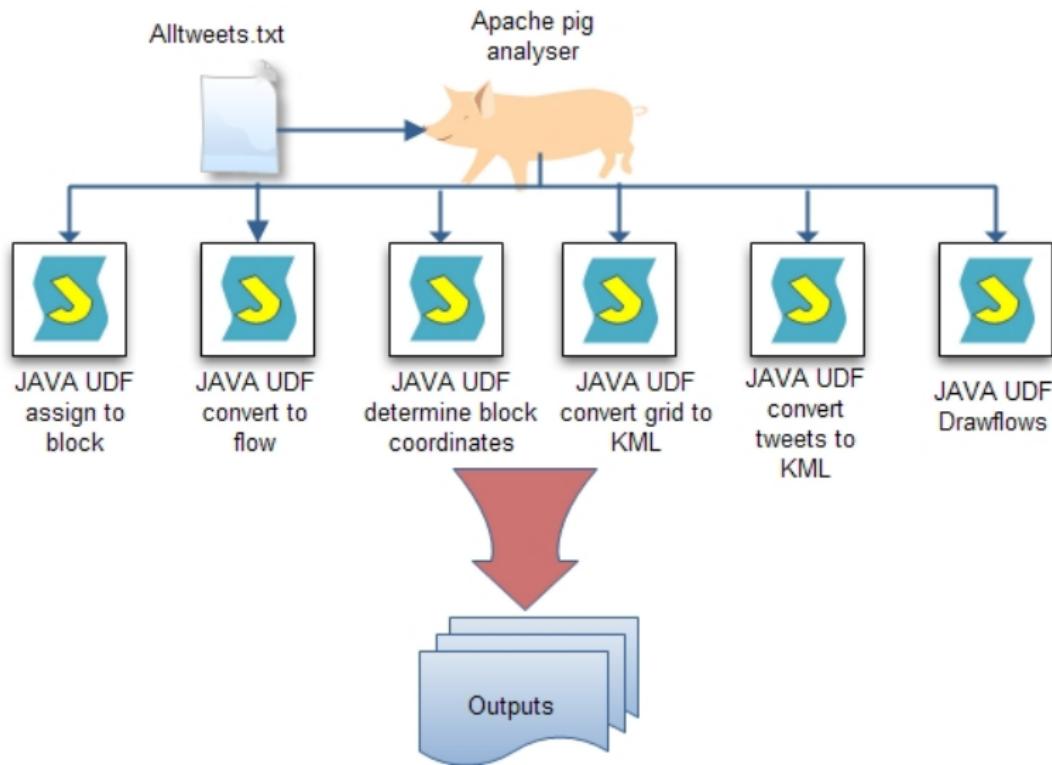
What?

- Divide the country in 4 blocks at a time
- Stop when an individual block is small enough
- I.e It has less than distinct 1000 users.

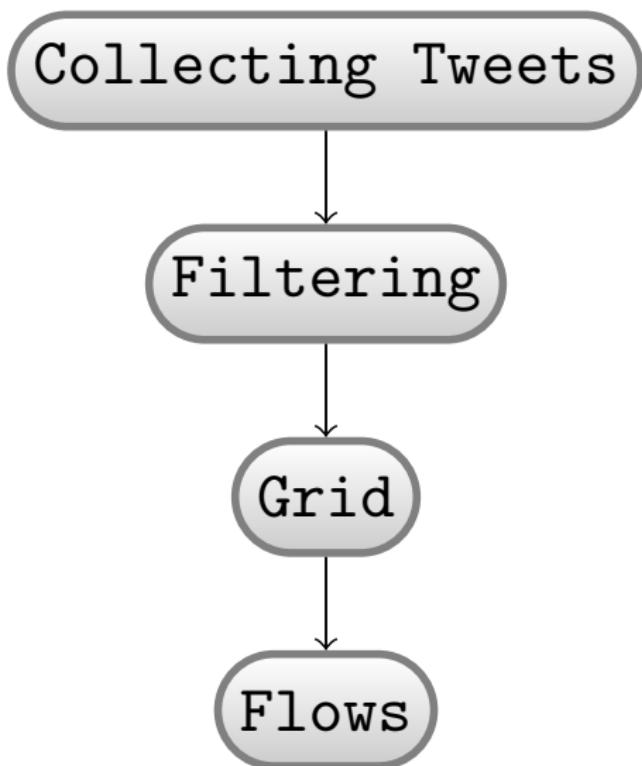
Why?

- To dynamically distinguish between population hubs
- To make blocks comparable in user base size
- To identify meaningful movements

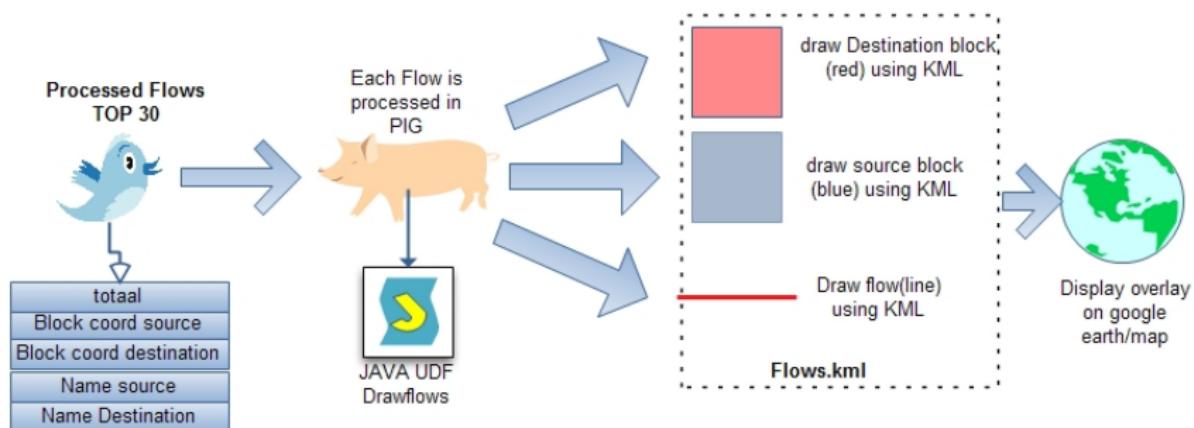
Pig



Advanced data processing



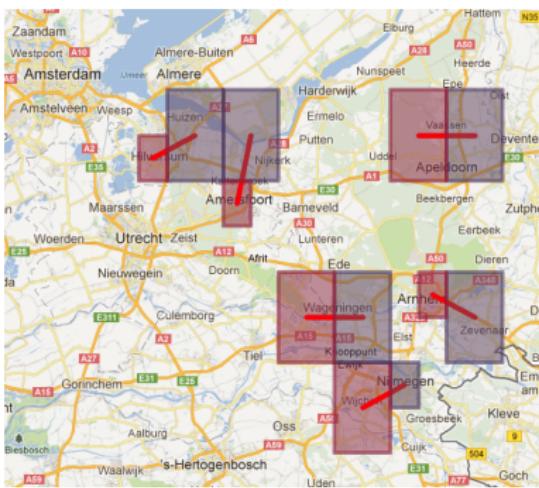
Presentation of the data |



Presentation of the data II



(a) Major nationwide flows



(b) Example of some flows

Geolocation of Flows

Correlation coefficient

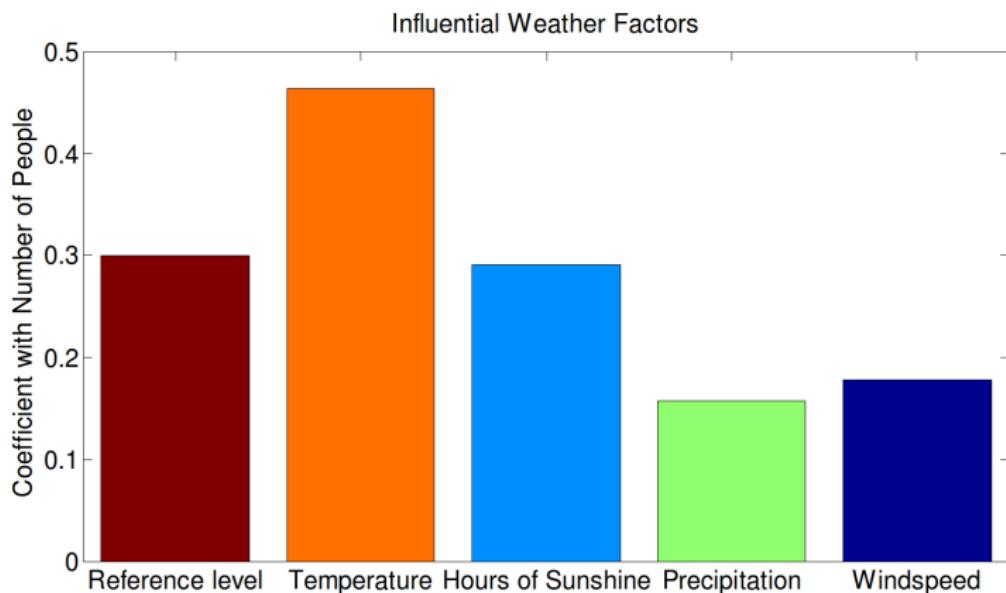
$$r_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}}$$

The correlation coefficient whose magnitude are less than **0.3** have little if any (linear) correlation

Coefficient for 8 stations

Station	Temperature	Hours of Sunshine	Precipitation	Mean Windspeed
257	-0.4341	0.368	-0.0718	NA
269	-0.6196	-0.1973	-0.1902	-0.1744
270	-0.452	0.1607	0.3867	0.3517
275	-0.6017	-0.2435	-0.0331	-0.0775
283	-0.5684	0.1562	-0.0582	-0.1737
350	-0.5932	0.2906	0.0916	-0.1816
375	-0.2337	0.5927	-0.1964	0.173
380	-0.2043	0.3126	-0.2323	-0.1093

Number of people compared to weather factors I



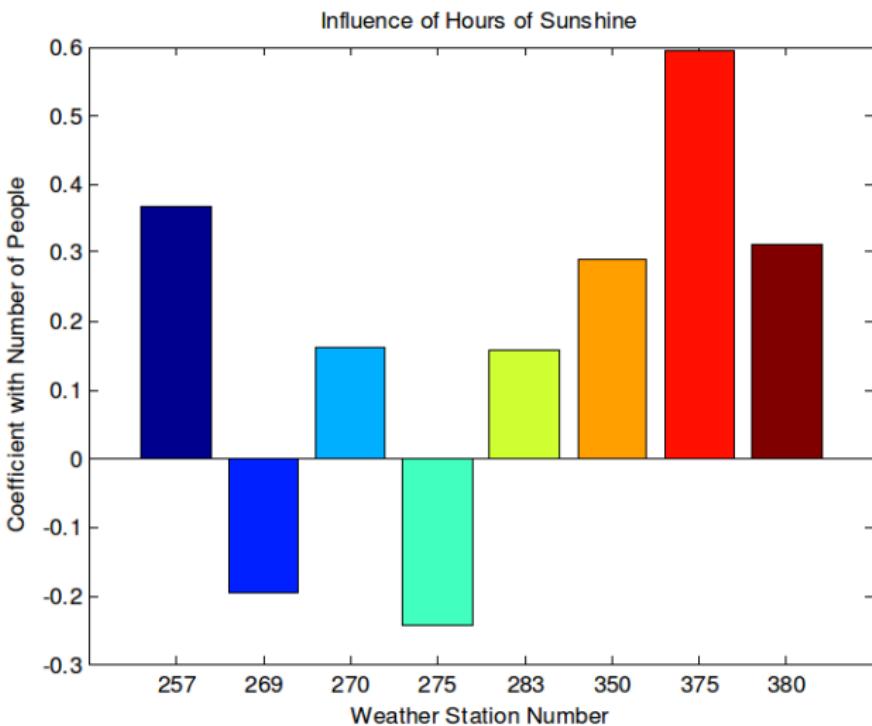
Number of people compared to weather factors II

- ① Only the factor of **temperature** is above the coefficient reference level (0.3), thus statistically the dominant weather factor that influences people's mobility.
 - ② Coefficients of **temperature** for all 8 stations are negative values. Colder weather encourages people to move more frequently.

Number of people compared to weather factors III

- ③ Coefficient of **precipitation** and **wind speed** is very low, thus have little influence. This may be due to the significant variation during the monitoring period.
 - ④ Coefficient of **Hours of sunshine** is close to the threshold. Clear trend shows that coefficient is positive value. Thus, more sunshine, more people tweeting.

Hours of sunshine from eight weather stations



Discussion

① Twitter

- Selected group of people : Target group
 - Other groups?
 - Other Social Media?
 - Other resources? (Questionnaire)

② Time

- Collection time: about one month
 - Slight variation of weather

③ **Solution:** Longer time period + combination of resources

Solution

- Clear correlation between temperature and amount of people tweeting in different locations (short distance)
 - Assumption: temperature and mobility are correlated
- Order of influence:
 - ① Temperature
 - ② Hours of Sunshine
- Nearly no influence:
 - Precipitation
 - Wind Speed
- Regional Differences

Video