

# DATA 609

Final Presentation

Dan Fanelli

Fall 2016

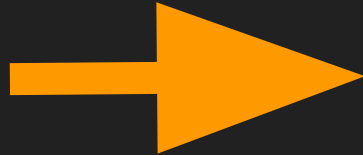
1. Social Network Analysis (Section 8.3 #1 - p322)
2. Intersection Data Simulation (Section 5.5 #3 - p222)

# Social Network Analysis

Section 8.3 #1 - p322

# Social Network Analysis

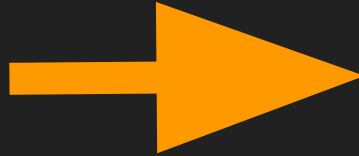
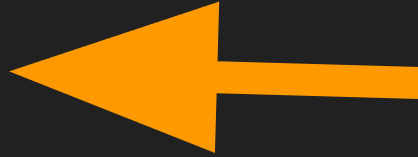
From Wikipedia's "List\_of\_social\_networking\_websites", GENI was chosen:



According to Wikipedia: "Geni is a commercial genealogy and social networking website" that "aimed to create a family tree of the world."

# Tools used for Geni Social Network Analysis

Hadley Wickam's rvest  
for data scraping



RNeo4j for  
persistence of social  
network information.

# Graph Design

- Each person in the ancestry graph was a VERTICE
- Each PARENT RELATION was an EDGE
- “Sideways” relations, ie. Siblings and other non-parent relations, were not scraped for the sake of model simplicity.
- The method was fed a ROOT person, and 3 generations of ancestry for that person were extracted (root + 3 levels of parents, grandparents, etc.)

# Run #1 GINI: Kevin Bacon



## Kevin Norwood Bacon MP

Gender: Male

Current Location: New York, NY, USA

Birth: July 8, 1958  
North Fayette, PA, USA

Immediate Family: Son of [Edmund Norwood Bacon](#) and [Ruth Hilda Holmes](#)  
Husband of [Kyra Sedgwick](#)  
Father of [<private> Bacon](#) and [<private> Bacon](#)  
Brother of [<private> Bacon](#); [<private> \(Bacon\)](#); [<private> \(Bacon\)](#); [<private> \(Bacon\)](#); [<private> \(Bacon\)](#); [<private> \(Bacon\)](#) and 2 others

Added by: [M](#) ❤ on June 24, 2009

Managed by: [Michael Reid Delahunt, PhD](#) and 4 others

Curated by: [Jenna, Volunteer Curator](#)

Overview

Media

Timeline

Discussions (1)

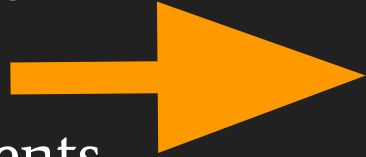
Sources

Revisions

DNA

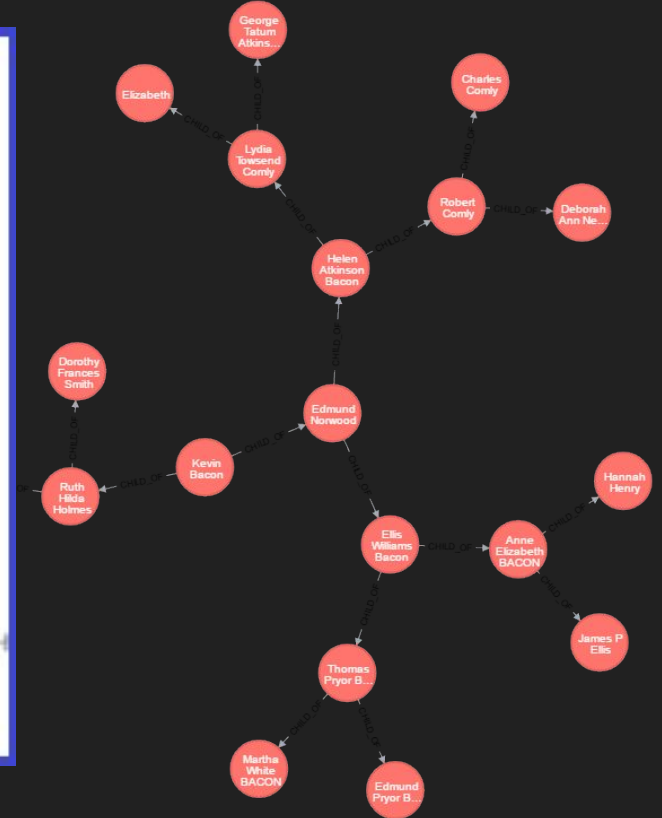
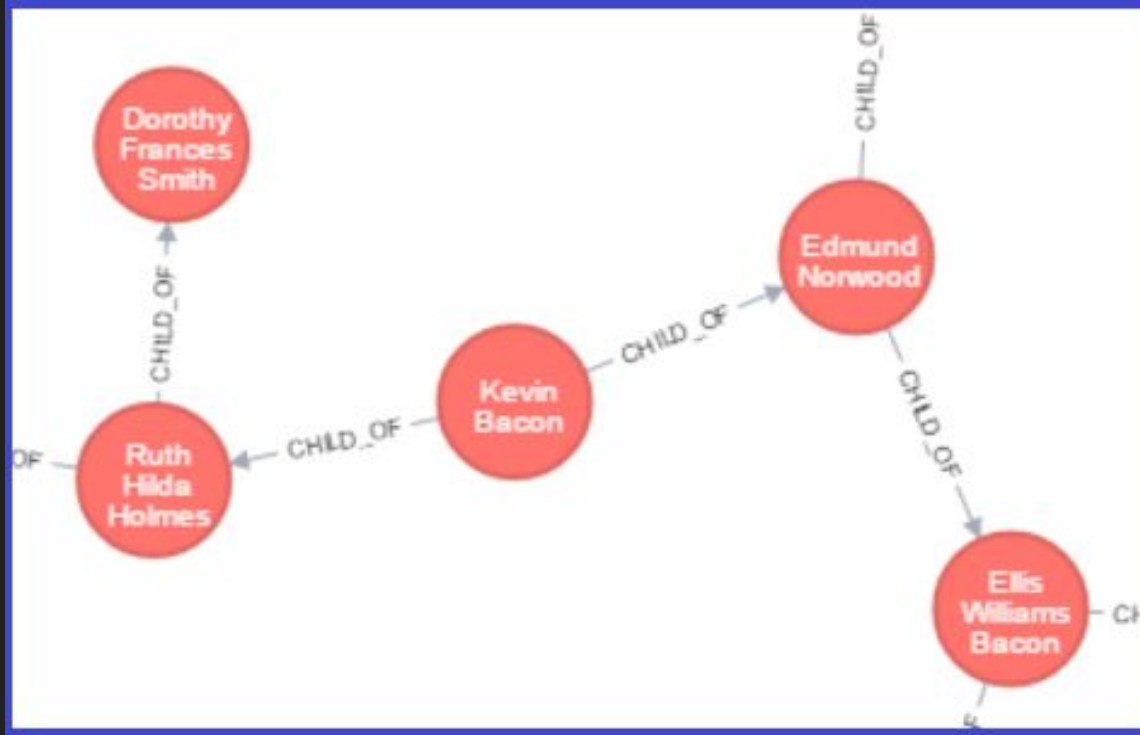
# Run #1 Printouts: Kevin Bacon

- 1 = Kevin Bacon
- 2 = Parents
- 3 = Grandparents
- 4 = Great  
Grandparents



```
## [ 1 ] Kevin Bacon
## [ 2 ] Edmund Norwood Bacon
## [ 3 ] Ellis Williams Bacon
## [ 4 ] Thomas Pryor BACON
## [ 4 ] Anne Elizabeth BACON
## [ 3 ] Helen Atkinson Bacon
## [ 4 ] Robert Comly
## [ 4 ] Lydia Townsend Comly
## [ 2 ] Ruth Hilda Holmes
## [ 3 ] Artemas Holmes
## [ 4 ] Artemus Henry Holmes
## [ 4 ] Lillian Holmes
## [ 3 ] Dorothy Frances Smith
## [ 4 ] George Campbell Smith
## [ 4 ] Annie Schwartz
```

# Run #1 Neo4j Graph: Kevin Bacon





# Run #2 GINI: JFK



## John Fitzgerald Kennedy, Sr. MP 100

Gender: Male

Birth: May 29, 1917  
Brookline, Norfolk, Massachusetts, United States

Death: November 22, 1963 (46)  
Dallas, Texas, United States (murdered)

Place of Burial: Arlington National Cemetery, Arlington, VA, USA

Immediate Family: Son of [Joseph P. Kennedy, Sr.](#) and [Rose Elizabeth Kennedy](#)  
Husband of [Jacqueline Kennedy Onassis](#)  
Father of [Arabella Kennedy](#); [Caroline Bouvier Schlossberg](#); [John F. Kennedy, Jr.](#) and [Patrick Bouvier Kennedy](#)  
Brother of [Joseph P. Kennedy, Jr.](#); [Rosemary Kennedy](#); [Kathleen, Marchioness of Hartington](#); [Eunice Mary Shriver](#); [Patricia Helen Lawford](#) and 3 others

Added by: [Rick](#) on June 2, 2007

Managed by: [Tina](#) and 51 others

Curated by: [Daniel Dupree Walton](#)

Overview

Media

Timeline


Discussions (3)

Sources (2)

Revisions

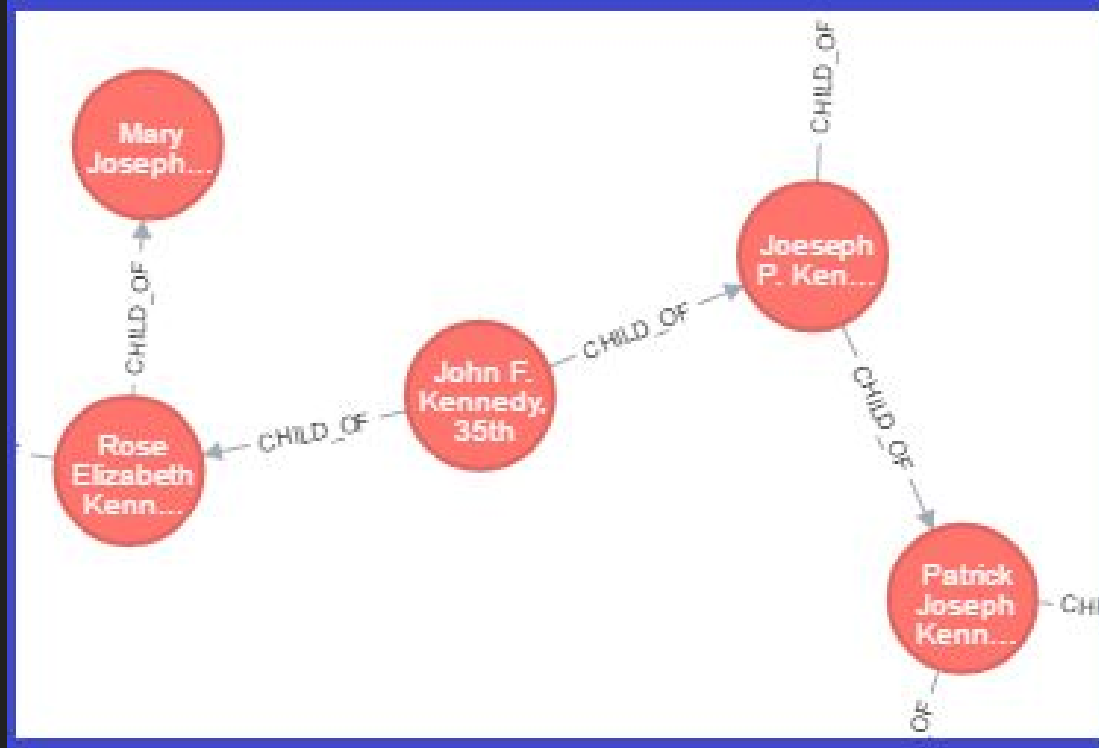
DNA

# Run #2 Printouts: JFK

- 1 = John F. Kennedy
  - 2 = Parents
  - 3 = Grandparents
  - 4 = Great Grandparents
- 

```
## [ 1 ] John F. Kennedy, 35th President of the USA
## [ 2 ] Joseph P. Kennedy, Sr.
## [ 3 ] Patrick Joseph Kennedy, Jr.
## [ 4 ] Patrick Joseph Kennedy, Sr.
## [ 4 ] Bridget Kennedy
## [ 3 ] Mary Augusta Kennedy
## [ 4 ] James Hickey
## [ 4 ] Margaret Hickey
## [ 2 ] Rose Elizabeth Kennedy
## [ 3 ] John F. "Honey Fitz" Fitzgerald, Sr.
## [ 4 ] Thomas Fitzgerald
## [ 4 ] Rosanna Fitzgerald
## [ 3 ] Mary Josephine Fitzgerald
## [ 4 ] Michael Hannon
## [ 4 ] Mary Ann Fitzgerald
```

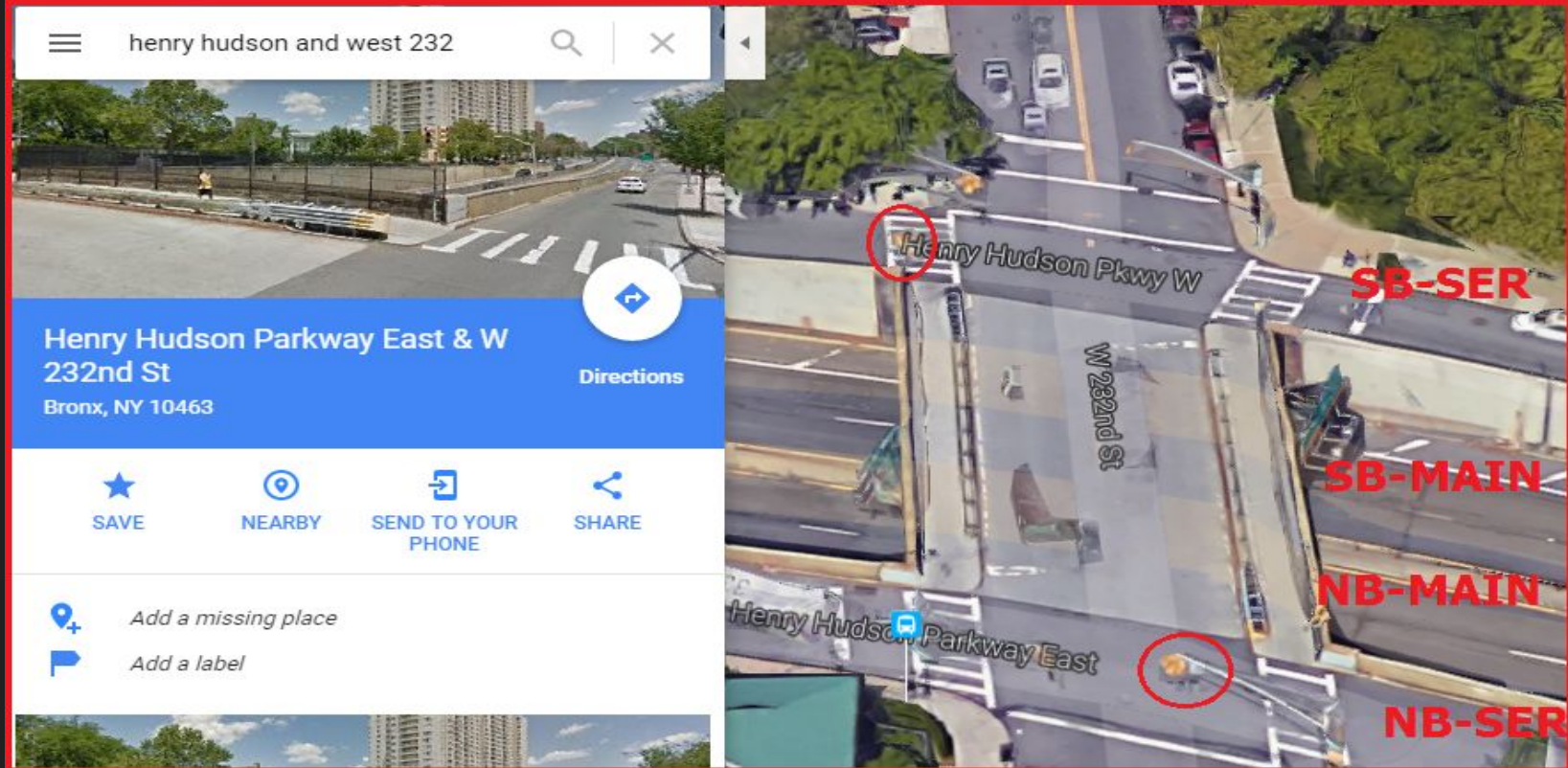
# Run #2 Neo4j Graph: JFK



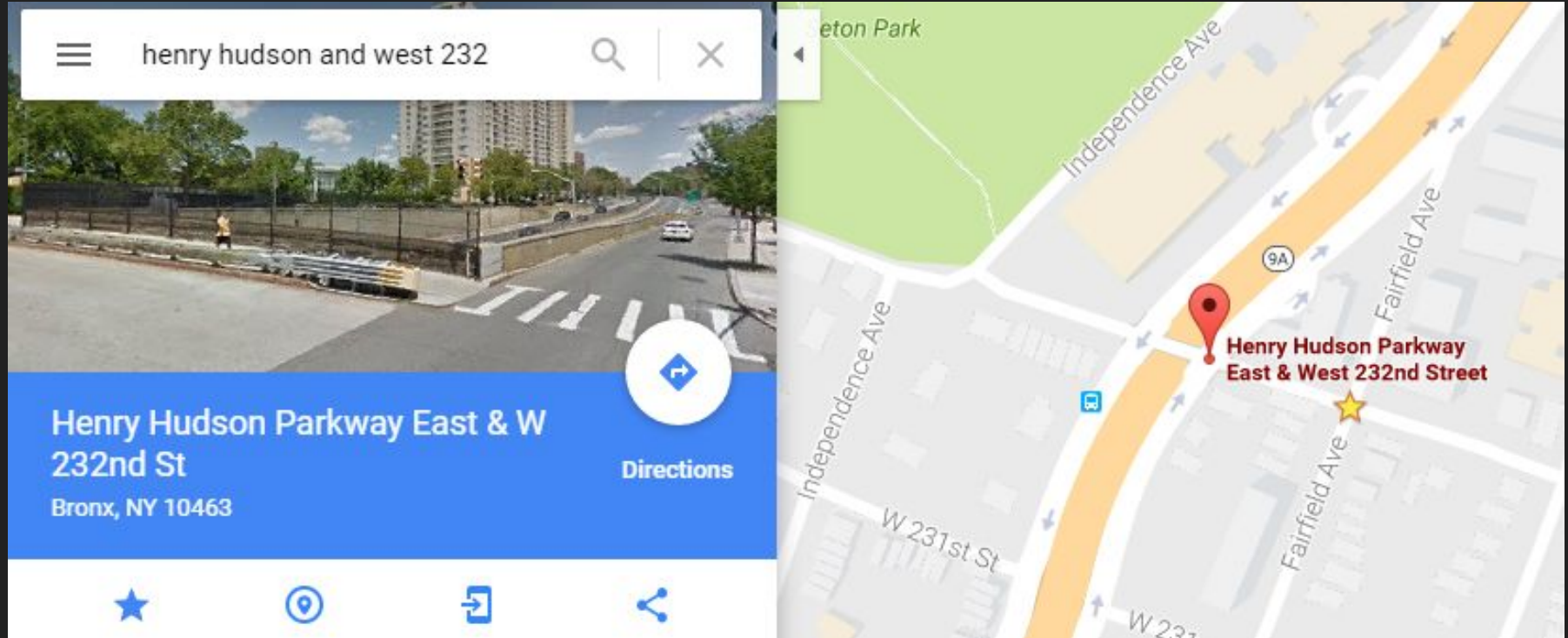
# Intersection Data Simulation

Section 5.5 #3 - p222

# Henry Hudson Pkwy and E 232nd St. (Arial)



# Henry Hudson Pkwy and E 232nd St. (Ground, Map)





# Traffic Data in original Wide Format

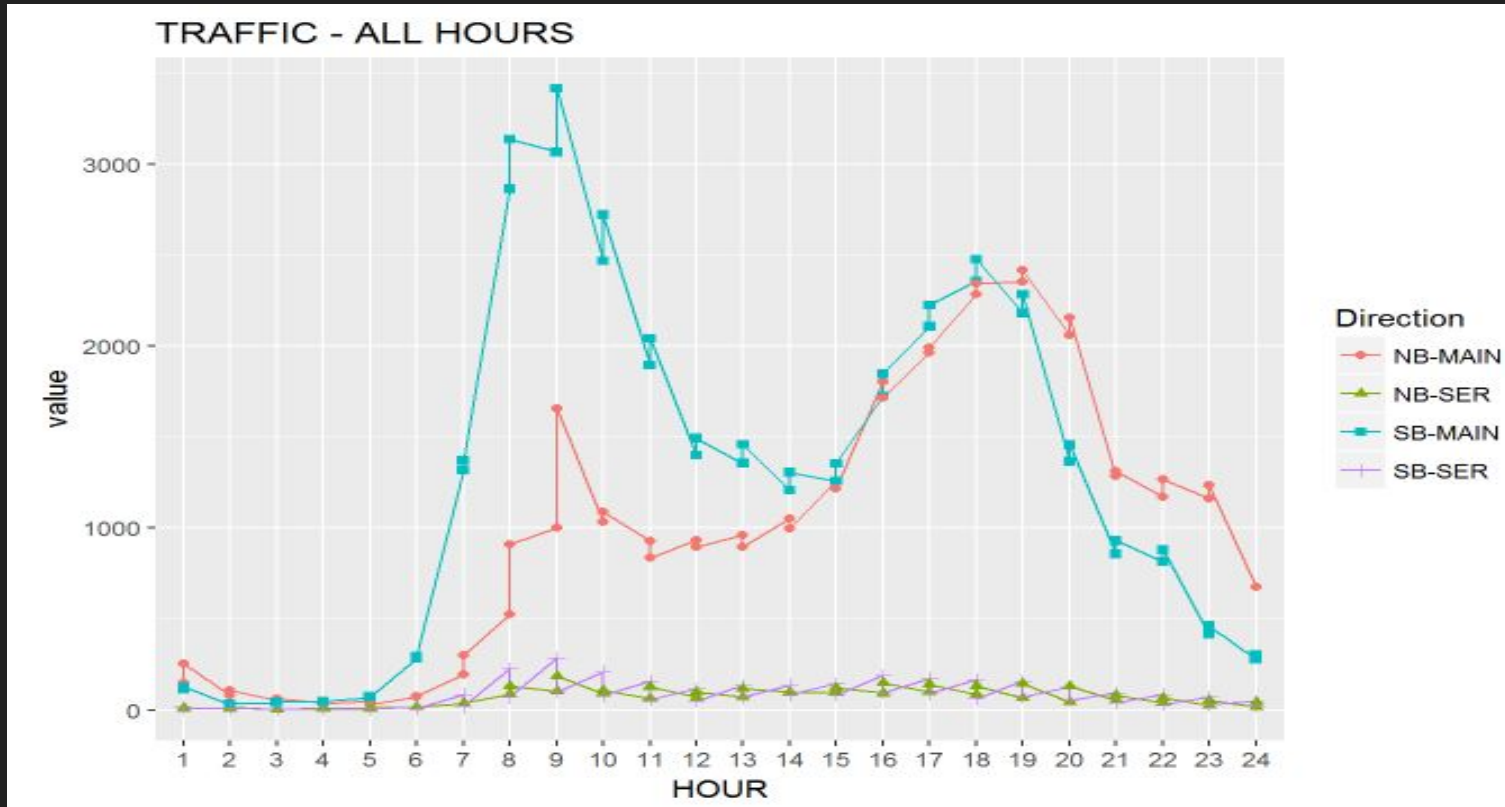
	ID	Segment_ID	Roadway_Name	xFrom	xTo	Direction	Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
5	361	174566	HENRY HUDSON PKWY	FAIRFIELD AVE	W 232nd ST	NB-MAIN	02/05/2013	150	81	51	38	46	71	195	526	1004	1035	930	934	960	1050	1258	1802	1963	22
12	361	174566	HENRY HUDSON PKWY	FAIRFIELD AVE	W 232nd ST	NB-SER	02/05/2013	5	8	4	11	5	17	35	87	104	90	62	69	71	99	97	92	98	
19	361	174566	HENRY HUDSON PKWY	FAIRFIELD AVE	W 232nd ST	SB-MAIN	02/05/2013	114	29	39	44	68	283	1317	2865	3067	2468	1896	1401	1357	1207	1259	1727	2108	23
26	361	174566	HENRY HUDSON PKWY	FAIRFIELD AVE	W 232nd ST	SB-SER	02/05/2013	20	9	4	4	9	15	87	226	286	210	157	116	135	136	143	190	174	1
33	251	73220	HENRY HUDSON PKWY	INDEPENDENCE AVE	W 232nd ST	NB-MAIN	02/05/2013	254	108	64	34	31	78	300	908	1656	1091	837	898	898	995	1217	1718	1990	23
40	251	73220	HENRY HUDSON PKWY	INDEPENDENCE AVE	W 232nd ST	NB-SER	02/05/2013	13	15	4	7	10	16	37	130	186	109	127	98	118	96	121	148	142	1
47	251	73220	HENRY HUDSON PKWY	INDEPENDENCE AVE	W 232nd	SB-MAIN	02/05/2013	126	33	42	46	73	297	1371	3137	3418	2723	2043	1494	1461	1306	1354	1848	2225	24

# Traffic Data “Reshaped” to Long Format

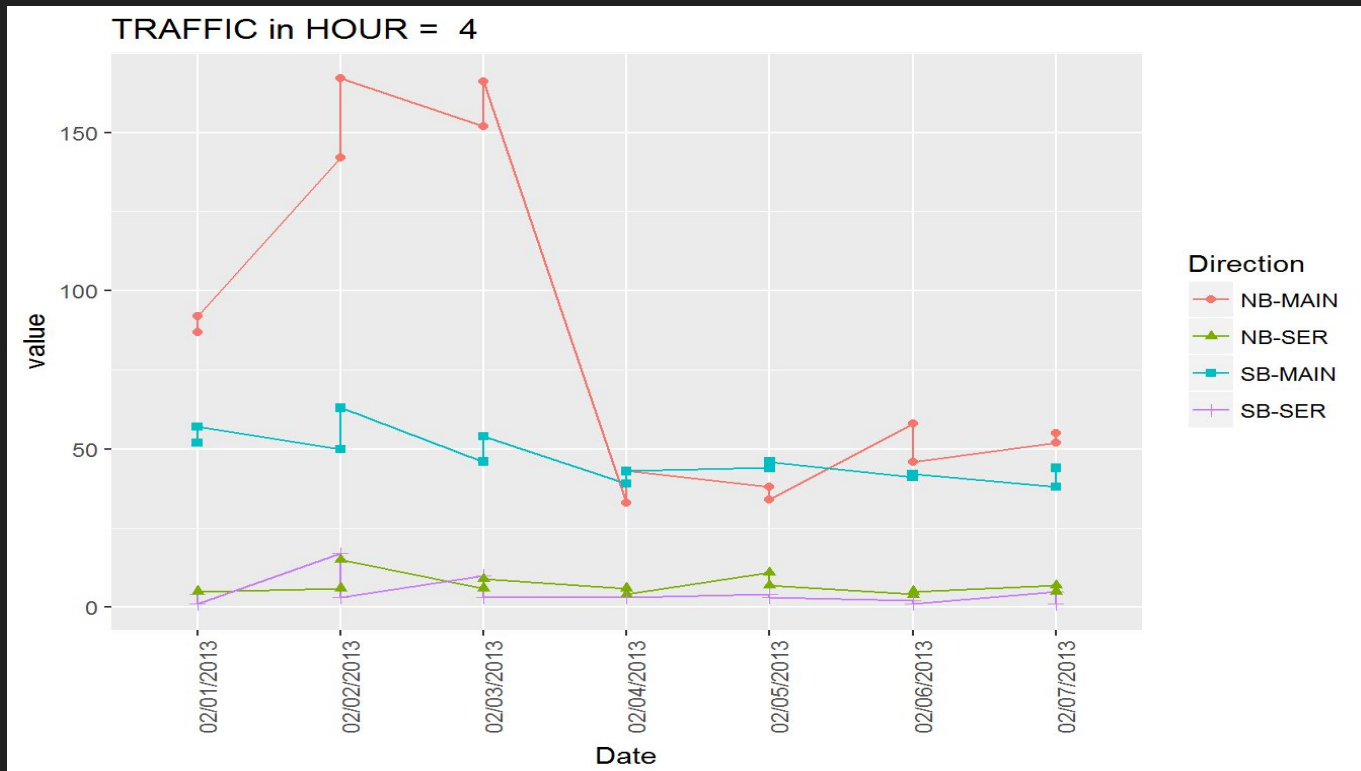
	ID	Segment_ID	Roadway_Name	xFrom	xTo	Direction	Date	HOUR	value
5	361	174566	HENRY HUDSON PKWY	FAIRFIELD AVE	W 232nd ST	NB-MAIN	02/05/2013	1	150
12	361	174566	HENRY HUDSON PKWY	FAIRFIELD AVE	W 232nd ST	NB-SER	02/05/2013	1	5
19	361	174566	HENRY HUDSON PKWY	FAIRFIELD AVE	W 232nd ST	SB-MAIN	02/05/2013	1	114
26	361	174566	HENRY HUDSON PKWY	FAIRFIELD AVE	W 232nd ST	SB-SER	02/05/2013	1	20
33	251	73220	HENRY HUDSON PKWY	INDEPENDENCE AVE	W 232nd ST	NB-MAIN	02/05/2013	1	254
40	251	73220	HENRY HUDSON PKWY	INDEPENDENCE AVE	W 232nd ST	NB-SER	02/05/2013	1	13
47	251	73220	HENRY HUDSON PKWY	INDEPENDENCE AVE	W 232nd ST	SB-MAIN	02/05/2013	1	126
54	251	73220	HENRY HUDSON PKWY	INDEPENDENCE AVE	W 232nd ST	SB-SER	02/05/2013	1	7
61	361	174566	HENRY HUDSON PKWY	FAIRFIELD AVE	W 232nd ST	NB-MAIN	02/05/2013	2	81



# HOURLY data for sample date: (02/05/2013)



# Maybe $\max(\text{NB}, \text{SB})$ in Hour 4 is Random?



# Aah, those are Weekend Days...



what day of week was Feb 2, 2013



**All**

Shopping

News

Videos

Maps

More

Settings

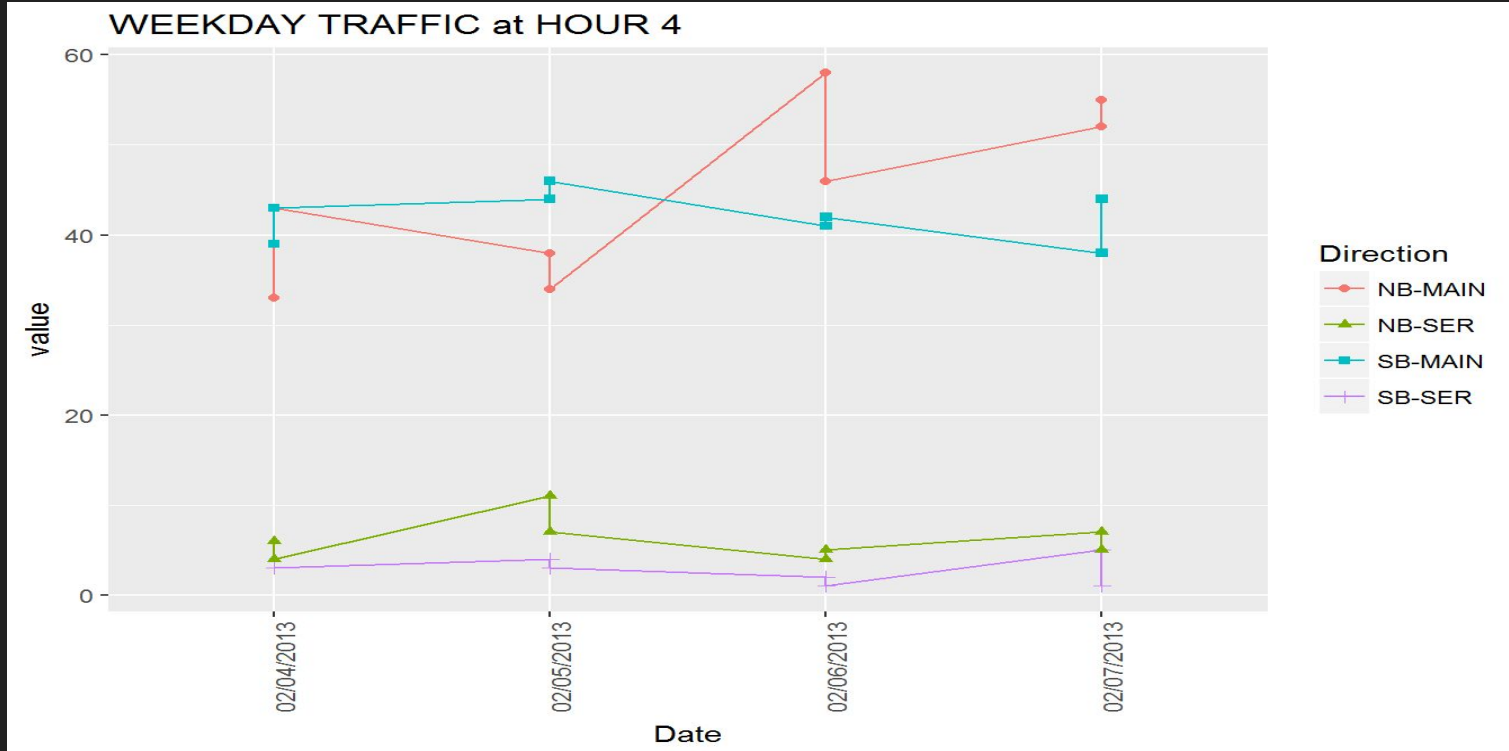
Tools

About 43,500,000 results (0.83 seconds)

## Saturday

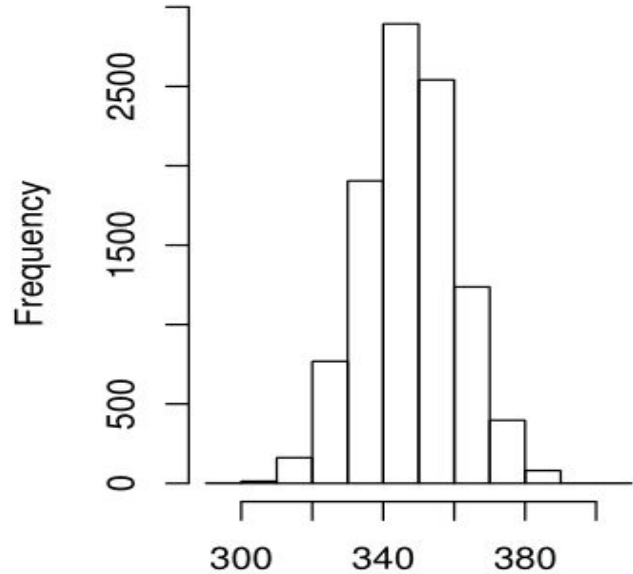
February 2, 2013

# Use only the final 4 Weekdays for Rush Hour



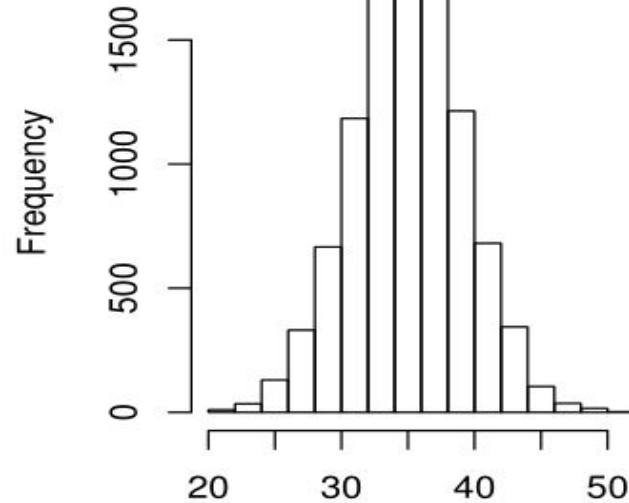
# Monte Carlo Simulation:

**4282 / 10000 over 4AM count**



NB-MAIN and SB-MAIN

**28 / 10000 over 4AM count**



NB-SER and SB-SER

## Conclusions:

- The probability that MAIN road fits null hypothesis (no direction bias) at hour 4 is:  
0.4274
- The probability that SERVICE road fits null hypothesis (no direction bias) at hour 4 is:  
0.0024