



# Data Visualization

Connecting research to the world

# Data Visualization

CONNECTING RESEARCH TO THE WORLD

## Why is visualizing data important?

A big problem with social research (and many other types of research) today is that it doesn't reach many of the most important audiences (the public, research participants etc.)

What does research generally produce?

**HUGE BORING REPORTS NOBODY WANTS TO READ!**



# Data Visualization

CONNECTING RESEARCH TO THE WORLD

## Why the disconnect?

- Most researchers communicate in language only researchers use
- Most organizations paying for research don't know to ask for something different
- Like all other industries, researchers have their own dogma – change is difficult

## Why the does it matter?

- Much of the research done is wasted
- People who are researched deserve to see the results
- The bigger the audience is, the more information you will draw from the research

# Data Visualization

CONNECTING RESEARCH TO THE WORLD

## Visualizing data is about telling a story

- Every data point has a story – just like a character in a book
- There are relationships, and interactions with data – just like characters in a book
- How you present data can help people remember
- Good visualization can make people take action



# Data Visualization

CONNECTING RESEARCH TO THE WORLD

## The art and science of data visualization

- Part scientist / researcher, part computer programmer, part graphic designer, part story teller

order	id	oblast	reason	number	interval	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2018	2041	2207	4 Ala Buika Butuk	12	Jumalshov	2	2	1	6	8	4	4	4	4	4	4	4	4	4	4
2018	2042	2201	4 Ala Buika Butuk	10	Jumalshov	3	1	1	7	8	4	4	4	4	4	4	4	4	4	4
2020	2051	2207	4 Ala Buika Butuk	22	Jumalshov	2	2	1	6	8	2	2	2	2	2	2	2	2	2	2
2023	2054	2210	4 Ala Buika Butuk	22	Jumalshov	2	1	1	6	8	4	4	4	4	4	4	4	4	4	4
2020	2060	2216	4 Ala Buika Butuk	31	Abdyrazzi	1	2	1	4	7	99	4	99	4	99	4	99	4	99	4
2020	2061	2217	4 Ala Buika Butuk	32	Abdyrazzi	2	1	1	4	7	4	4	4	4	4	4	4	4	4	4
2021	2062	2218	4 Ala Buika Butuk	33	Abdyrazzi	3	1	1	4	7	3	4	4	4	4	4	4	4	4	4
2022	2063	2219	4 Ala Buika Butuk	34	Abdyrazzi	2	1	2	3	7	4	4	4	4	4	4	4	4	4	4
2023	2064	2220	4 Ala Buika Butuk	35	Abdyrazzi	2	1	1	2	6	4	3	3	3	3	3	3	3	3	3
2024	2065	2221	4 Ala Buika Butuk	36	Abdyrazzi	2	1	1	2	6	4	4	4	4	4	4	4	4	4	4
2020	2066	2222	4 Ala Buika Butuk	37	Abdyrazzi	1	1	1	4	7	4	3	4	3	4	3	4	3	4	3
2026	2067	2223	4 Ala Buika Butuk	38	Abdyrazzi	2	1	1	4	7	4	3	99	4	3	99	4	3	99	4
2027	2068	2224	4 Ala Buika Butuk	39	Abdyrazzi	2	1	1	4	6	99	99	4	3	99	4	3	99	4	3
2028	2069	2225	4 Ala Buika Butuk	40	Abdyrazzi	1	1	2	3	4	4	4	4	4	4	4	4	4	4	4
2029	2070	2226	4 Ala Buika Butuk	41	Abdyrazzi	2	1	1	4	7	4	4	3	4	3	4	3	4	3	4
2040	2071	2227	4 Ala Buika Butuk	42	Abdyrazzi	2	1	2	4	5	4	4	4	4	4	4	4	4	4	4
2041	2072	2228	4 Ala Buika Butuk	43	Abdyrazzi	2	1	1	7	8	1	3	3	3	3	3	3	3	3	3
2042	2073	2229	4 Ala Buika Butuk	44	Abdyrazzi	2	1	2	3	6	4	3	3	3	3	3	3	3	3	3
2043	2074	2230	4 Ala Buika Butuk	45	Abdyrazzi	2	1	1	3	7	1	4	4	4	4	4	4	4	4	4
2044	2075	2231	4 Ala Buika Butuk	46	Abdyrazzi	2	1	1	2	5	4	4	4	4	4	4	4	4	4	4
2045	2076	2232	4 Ala Buika Butuk	47	Abdyrazzi	2	1	1	4	8	4	4	4	4	4	4	4	4	4	4
2046	2077	2233	4 Ala Buika Butuk	48	Abdyrazzi	2	1	1	3	6	4	4	4	4	4	4	4	4	4	4
2047	2078	2234	4 Ala Buika Butuk	49	Abdyrazzi	2	1	1	4	4	4	4	4	4	4	4	4	4	4	4
2048	2079	2235	4 Ala Buika Butuk	50	Abdyrazzi	2	1	1	4	5	4	4	4	4	4	4	4	4	4	4

Raw data



Computer programming



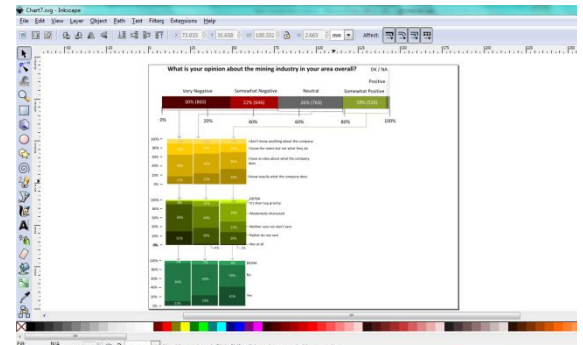
Graphic design

```
import matplotlib
matplotlib.use('Agg')
from pylab import *
import calendar

def webshow(img):
    savefig(img,dpi=500)
    print 'Content-Type: text/html\n'
    print '<img width="800" height="400" src="'+img+' />'

genres = []
n = 0
for c in sorted_list:
    genres.append(sorted_list[n][0])
    n += 1

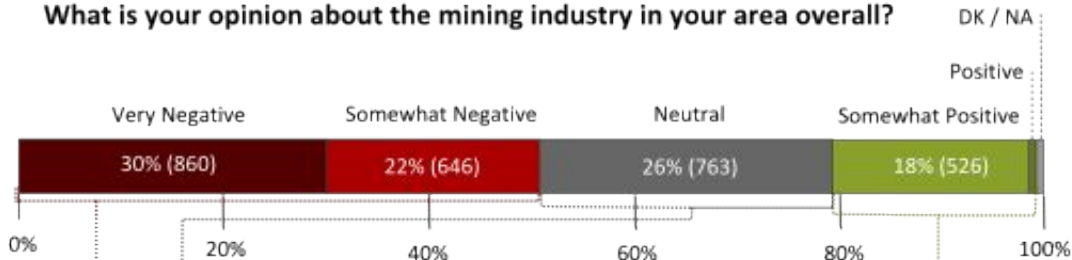
grosses = []
a = 0
for c in sorted_list:
    grosses.append(sorted_list[a][1])
```



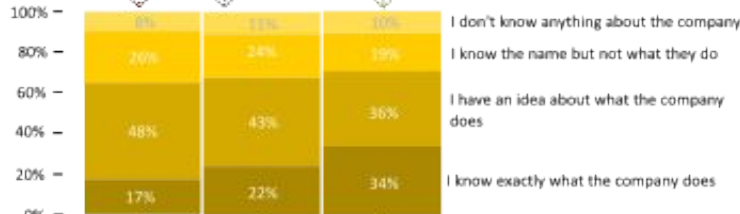
# Data Visualization

CONNECTING RESEARCH TO THE WORLD

What is your opinion about the mining industry in your area overall?

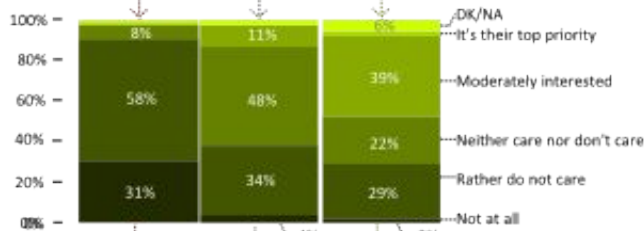


What do you know about the mining company in your rayon?



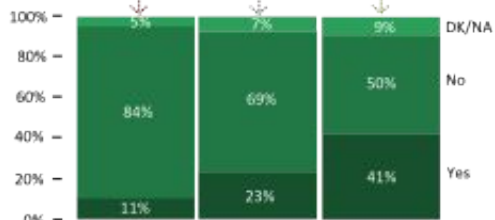
Those who viewed the industry positively were significantly more likely to claim exact knowledge of what the company in their area does. The precise cause of this relationship cannot be assessed from this data, however, it appears that greater knowledge of what the company is doing, may lead to a more positive assessment of the industry as a whole.

How much do you think mining companies care about providing benefits to your community?



Unsurprisingly, there is a direct link between perceptions of the overall industry and perceptions of how much companies care about providing benefits. Those with a negative view of the industry felt companies cared little or not at all about providing benefits.

Is your voice heard by the mining companies operating near your community?



This is one of the most important relationships examined. While a minority of people in all categories felt their voices were heard by companies in their area, those with a positive view of the industry were far more likely to feel their voices were heard.

# Data Visualization

CONNECTING RESEARCH TO THE WORLD

## Common types of “out-of-the-box” visualizations

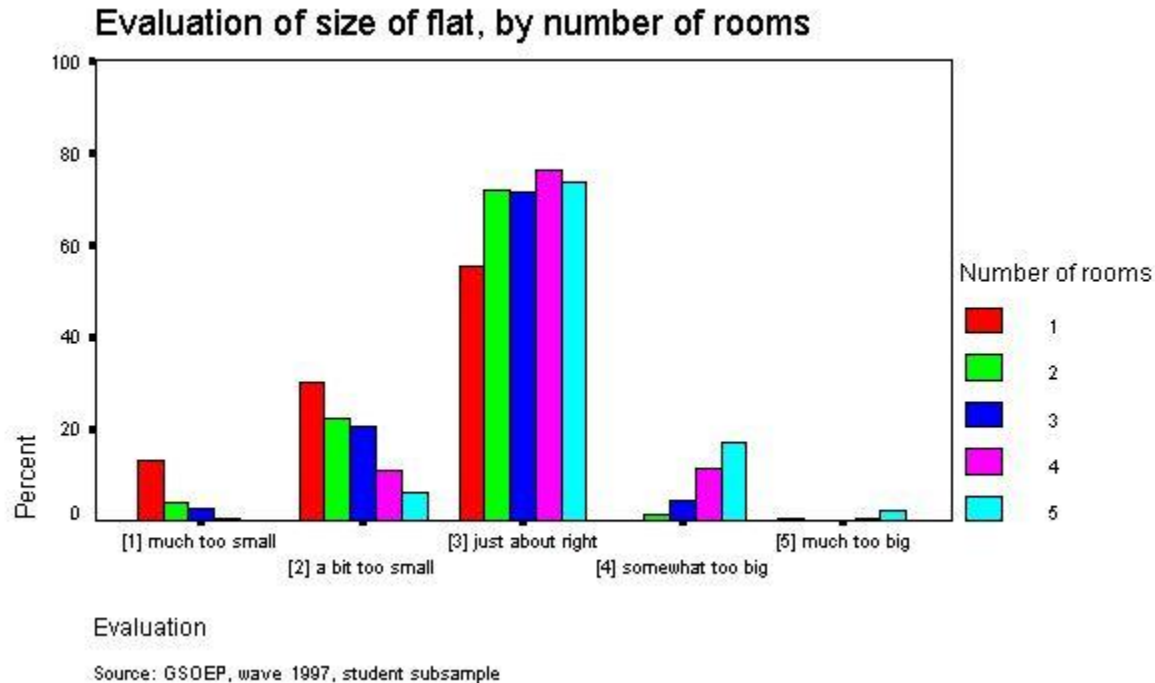
- Sociologist generally use SPSS
- Many other analysts use Excel
- Statisticians use “R”
- Graphic designers use Illustrator or Inkscape

All have their advantages and disadvantages as seen next...

# Data Visualization

CONNECTING RESEARCH TO THE WORLD

SPSS:



Advantages:

- Widely used in social sciences
- Extensive number of features
- Very rigorous analytical capabilities
- Able to handle large amounts of data

Disadvantages:

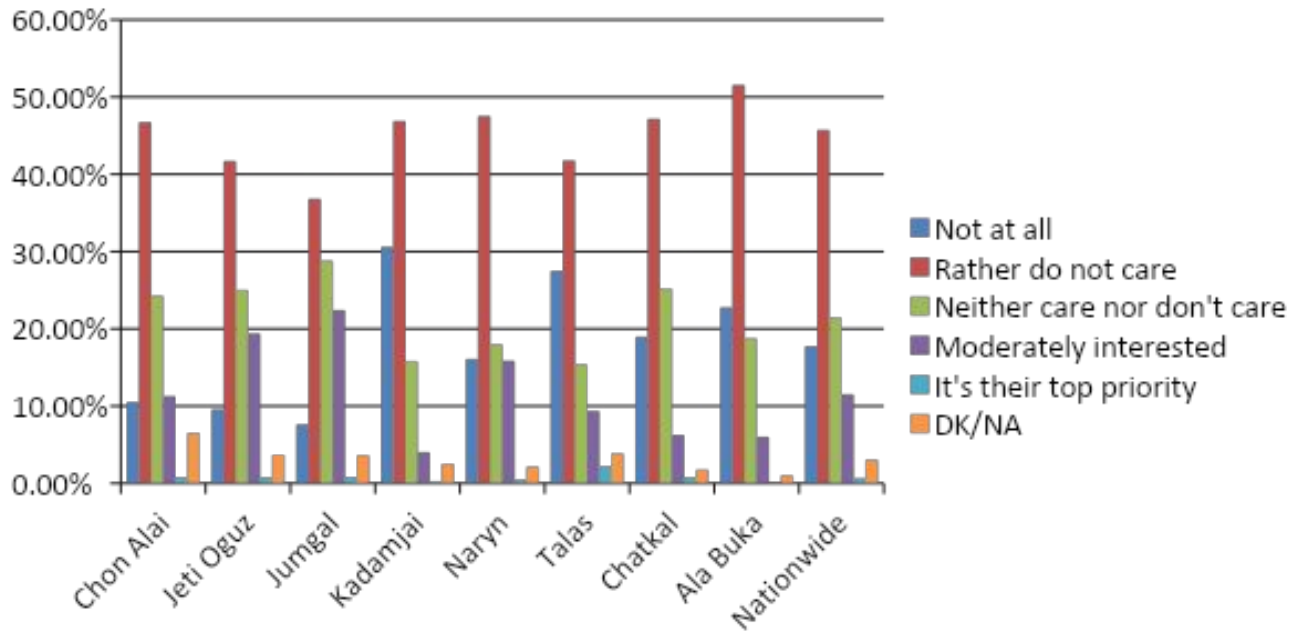
- UGLY charts
- Limited number of visualizations
- Incredibly expensive
- Not used by anyone but sociologists



# Data Visualization

CONNECTING RESEARCH TO THE WORLD

## Excel:



### Advantages:

- Used in nearly every office
- Very easy
- Able to handle somewhat large datasets
- Decent graphic quality

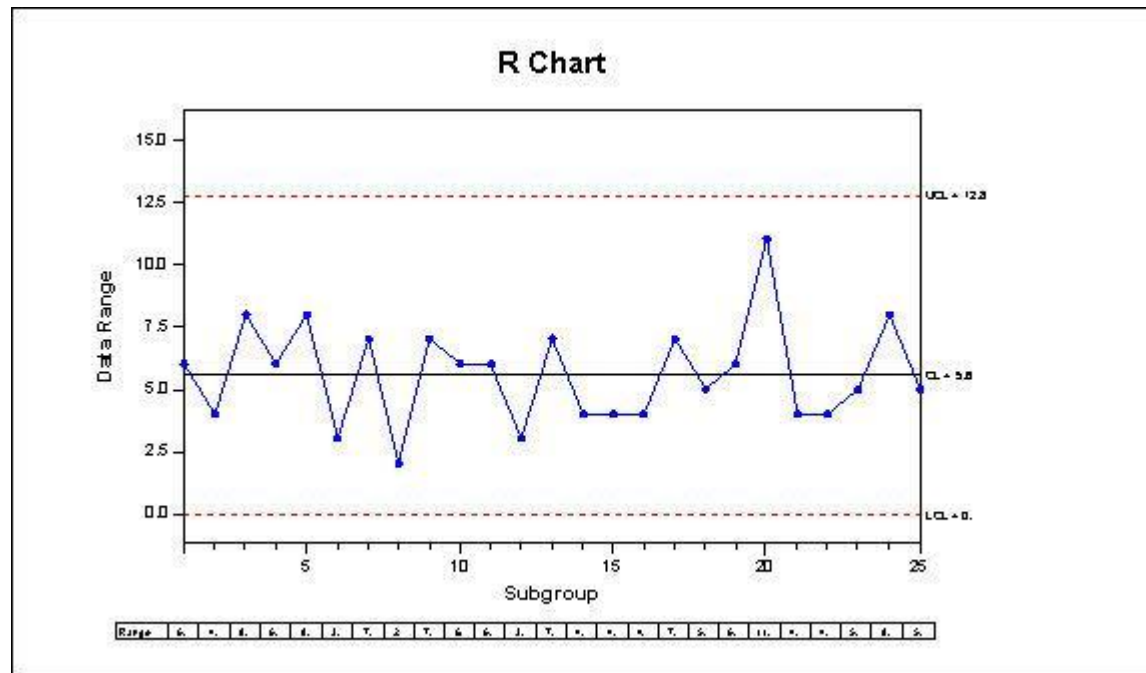
### Disadvantages:

- Limited chart types
- Not the best graphics

# Data Visualization

CONNECTING RESEARCH TO THE WORLD

R:



Advantages:

- Free! (legally)
- Able to handle large datasets
- Countless analytical functions

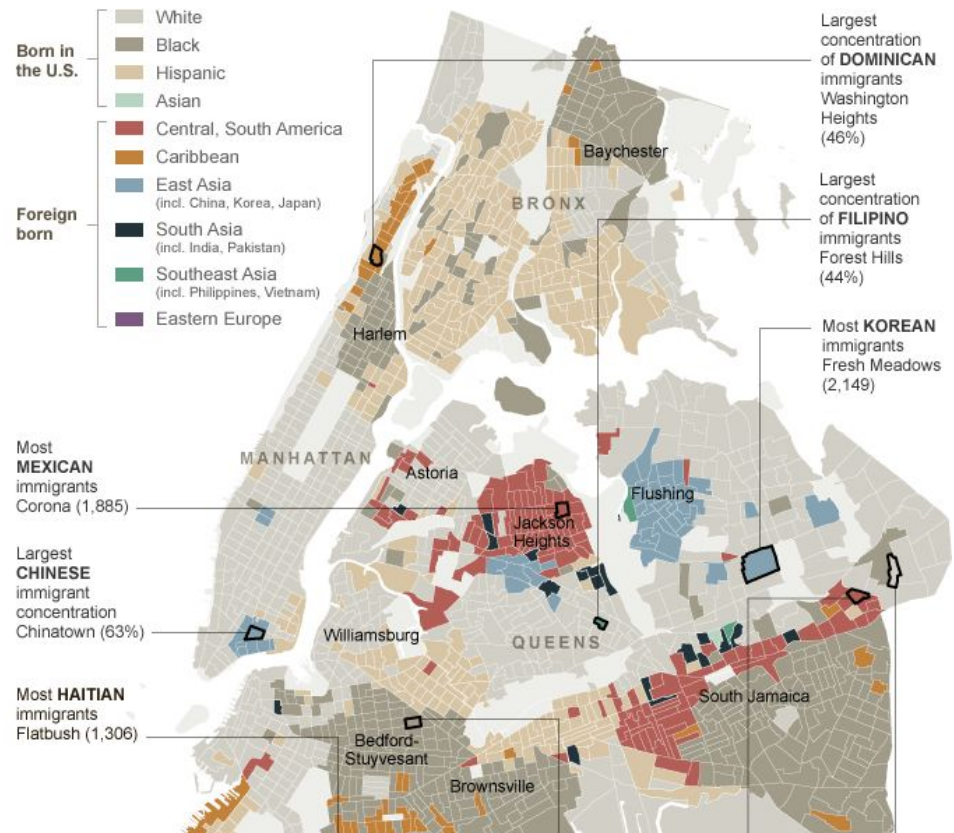
Disadvantages:

- Requires coding knowledge
- Ugly graphics

# Data Visualization

CONNECTING RESEARCH TO THE WORLD

Illustrator / Inkscape:



## Advantages:

- Amazing graphic quality
- Limitless flexibility for types of graphics

## Disadvantages:

- No analysis functions
- Easier to make mistakes

# Data Visualization

CONNECTING RESEARCH TO THE WORLD

The best outcomes are a mix of programming and graphic design

## Programming Options:

- Java
- Javascript
- HTML
- Python
- PHP
- Flash
- HTML

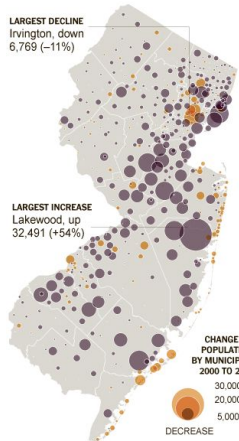
# Data Visualization

## CONNECTING RESEARCH TO THE WORLD

### Static examples:

**Total population** declined in some urban cores, but not in Newark, and grew in places like Lakewood, Franklin and Hoboken.

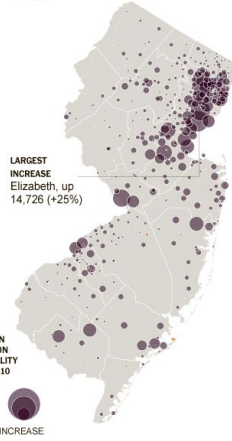
**LARGEST DECLINE**  
Irvington, down 6,769 (-11%)



**LARGEST INCREASE**  
Lakewood, up 32,491 (+64%)

**Hispanic** population grew widely. In inner cities like Paterson, Hispanics replaced blacks. In some smaller places, their numbers doubled.

**LARGEST INCREASE**  
Elizabeth, up 14,726 (+25%)

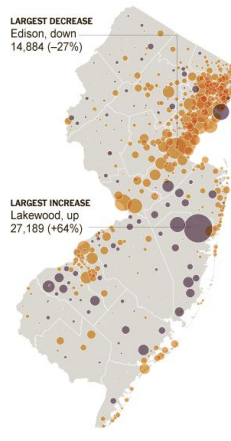


**CHANGE IN POPULATION BY MUNICIPALITY 2000 TO 2010**

30,000  
20,000  
5,000  
DECREASE INCREASE

**White** population fell across the state, giving way to Asians in Woodbridge and Hispanics in many places. It grew in Ocean County.

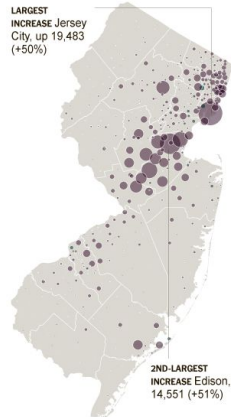
**LARGEST DECREASE**  
Edison, down 14,884 (-27%)



**LARGEST INCREASE**  
Lakewood, up 27,189 (+64%)

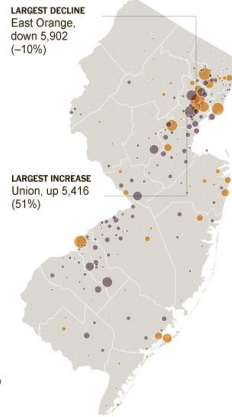
**Asian** population grew in most places, but the greatest increases were along Interstate 95, which also saw declines in white population.

**LARGEST INCREASE**  
Jersey City, up 19,483 (+50%)



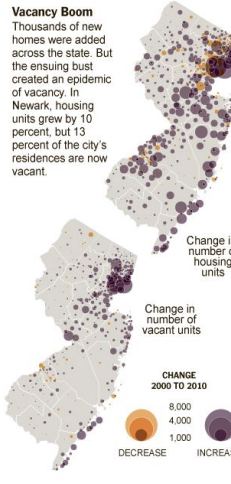
**Black** population declined in cities like Paterson, where Hispanic numbers grew, and increased in nearby areas.

**LARGEST DECLINE**  
East Orange, down 5,902 (-10%)

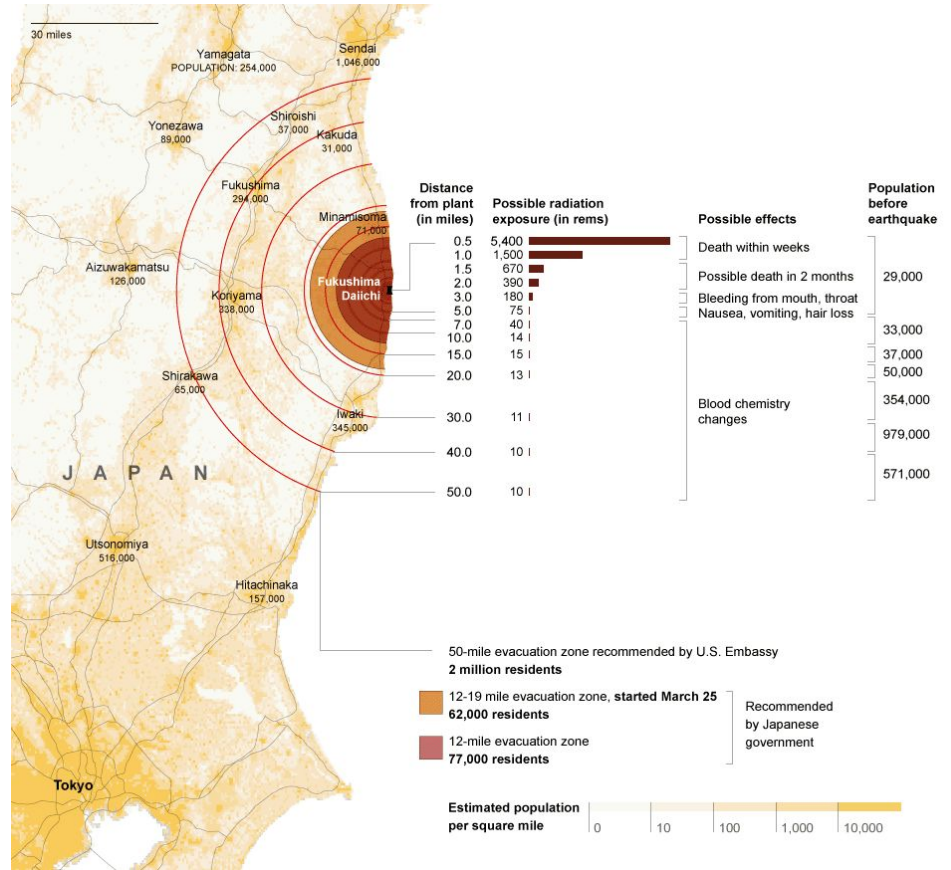


**LARGEST INCREASE**  
Union, up 5,416 (51%)

**Vacancy Boom**  
Thousands of new homes were added across the state. But the ensuing bust created an epidemic of vacancy. In Newark, housing units grew by 10 percent, but 13 percent of the city's residences are now vacant.



**CHANGE 2000 TO 2010**  
8,000  
4,000  
1,000  
DECREASE INCREASE





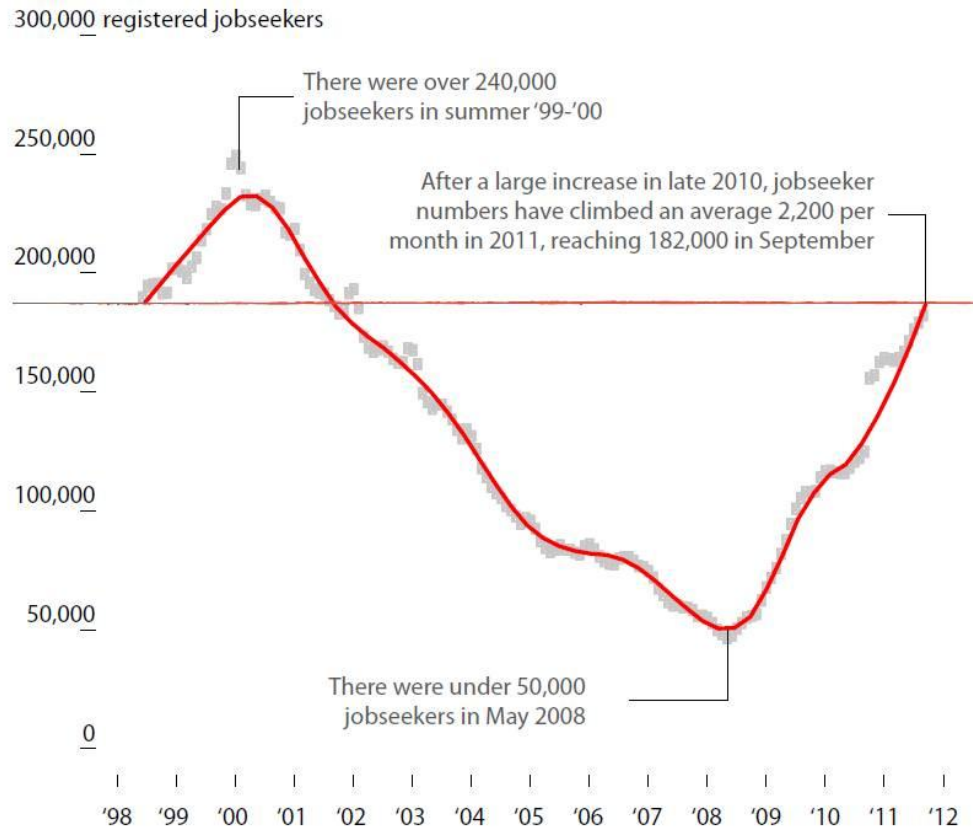
# Data Visualization

CONNECTING RESEARCH TO THE WORLD

## Static examples:

### NZ Unemployment Continues to Climb

Work and Income New Zealand greeted 18,000 new jobseekers between January and September this year, bringing the total number of jobseekers to 182,000. The last time unemployment reached this level was 1998-2001, when it continued to climb to almost 250,000.



# Data Visualization

CONNECTING RESEARCH TO THE WORLD

Dynamic examples: Tableau and Gapminder

<http://www.tableausoftware.com/learn/gallery/high-school-reading-and-math-scores>

<http://www.gapminder.org/world/>

<http://www.gapminder.org/videos/200-years-that-changed-the-world-bbc/>





# Data Visualization

What story do you want to tell?