Introductions

Þorri   **Palli**   **Matt**

Þorri: thorlakurl@ru.is
Palli: pallhilm@ru.is
Matt: matt@ru.is

Objectives

On completion of the course you should be able to:

- Take raw data and explain it to your audience
  using charts and maps
- Apply manual and automatic methods to collect data
- Apply manual and automatic methods to clean malformed data

This is a hands-on course in Data Acquisition and Visualization. It covers the entire process from finding and fetching (or integrating) data sources, through clean-up of data, to best-practices in visualizing the data in order to provide users with new insights or understanding. The concepts studied are applied in a practical project where the student applies the new knowledge to make a static or interactive visualization of data. The project concept and framework (data sources, selection of tools, etc.) are developed incrementally during the first two weeks, alongside lectures on topics to be applied; then combined into the outcome visualization in the last week which is devoted largely to project work.

Course Assessment

- **Question paper (10%)**
- **Presentation**
    - Concept (8%)
    - Implementation and methods (16%)
    - Quality of visualisations (32%)
    - Quality and content of the presentation (24%)
- **Report (10%)**

The grade will be based primarily on the visualization project, both intermediate deliverables and the final visualization. The course will end with a 5-10 minute presentation of the visualization deliverable, which will examine the understanding of the topics, the student's approach, methods and rationale behind individual project decisions.

Visualization Project

- **1-2 students per project**
- **Deliverables:**
  - **Final product can be "anything"**
    - Static images
    - Videos
    - Web page/site
    - Interactives
  - **Report summarise the data analysis and visualisation**
  - **5-10 minute presentations**
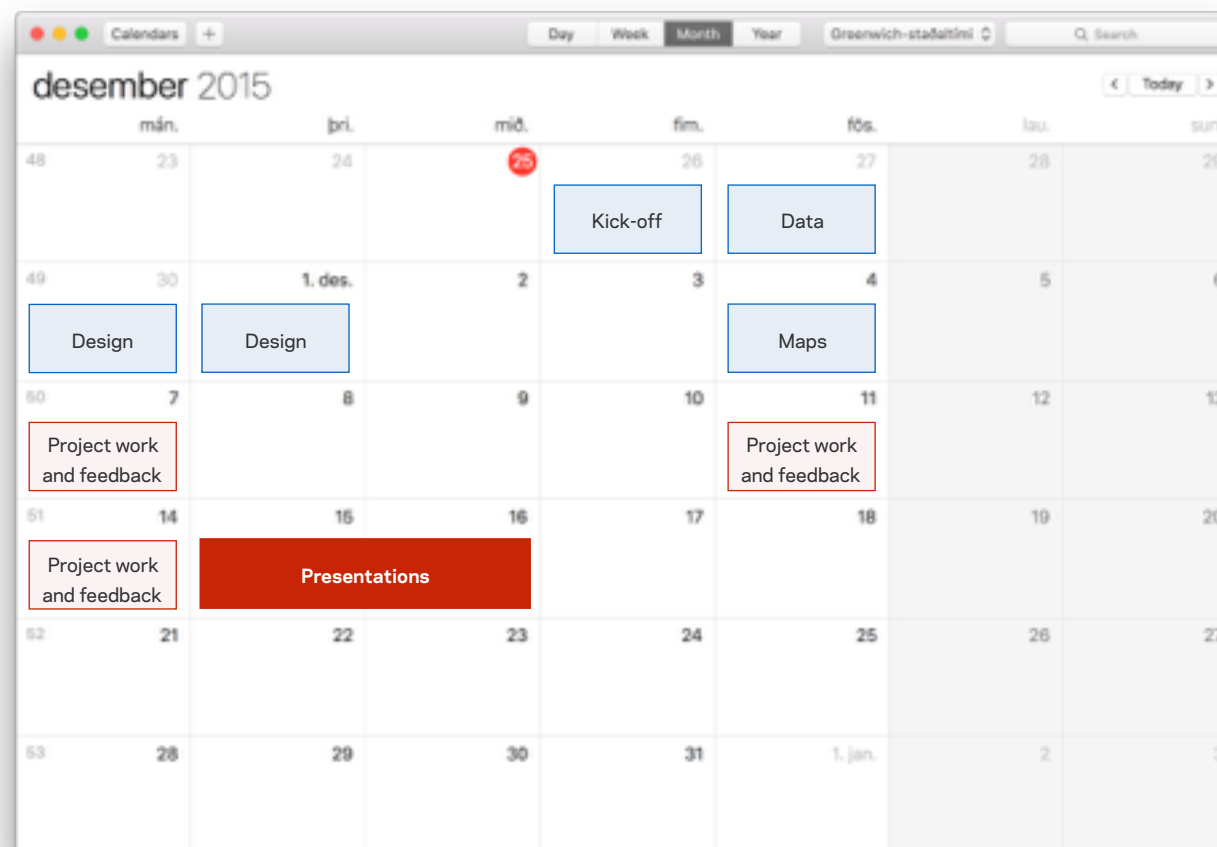
The Process

*„I keep saying that the sexy job in the next ten years will be statisticians"*

— Hal Varian

DATA
= NUMBERS & SYMBOLS
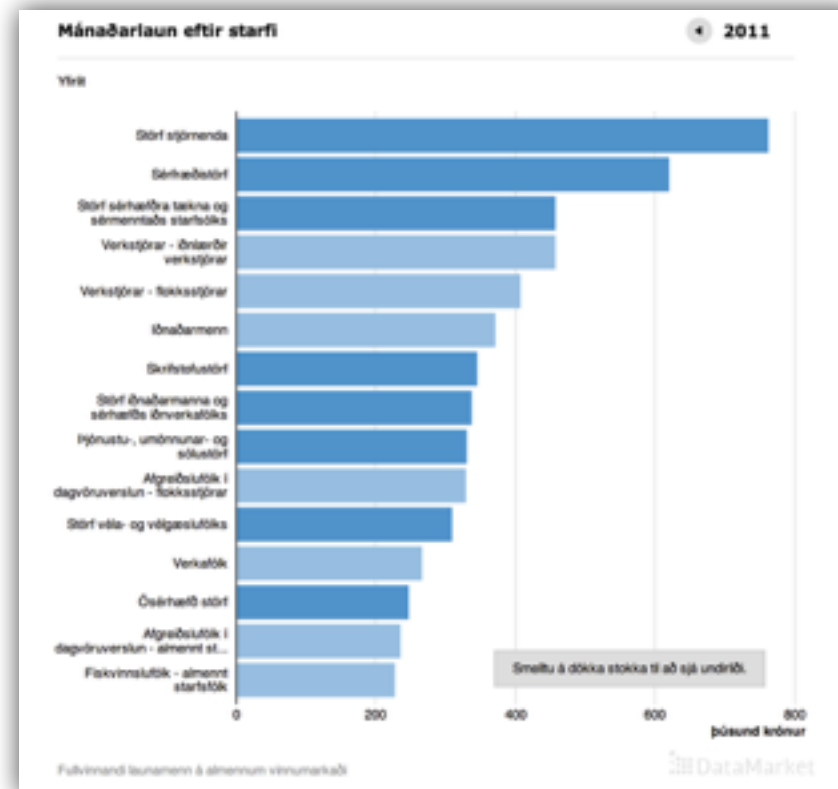
↓

INFORMATION
= DATA, PROCESSED for understanding

Mannfjöldi eftir kyni og aldri 1841-2008

	Á 1. ári		1 árs		2 ára		3 ára		4 ára		5 ára		6 ára		7 ára		
	Karlar	Konur	Karlar	Konur	Karlar	Konur	Karlar	Konur	Karlar	Konur	Karlar	Konur	Karlar	Konur	Karlar	Konur	Karl
1841	968	903	587	594	569	539	608	603	668	708	652	678	673	740	695	762	8?
1842	993	994	671	654	568	576	567	534	607	600	665	704	649	675	669	737	6?
1843	1.025	1.013	723	762	659	640	564	574	566	532	604	597	662	701	647	673	6?
1844	795	807	719	752	702	739	657	636	563	571	561	529	601	594	659	699	6?
1845	859	859	641	672	714	743	700	735	655	634	560	569	559	526	598	593	6?
1846	890	834	725	766	635	664	709	740	696	733	655	634	559	568	558	525	5?
1847	613	604	681	621	642	679	629	656	705	733	695	731	652	630	556	565	5?
1848	706	696	539	502	666	608	635	674	625	653	701	732	691	731	650	627	5?
1849	870	936	613	583	526	490	658	605	632	672	621	651	699	731	689	731	6?
1850	890	875	705	828	585	550	503	465	640	587	621	665	613	644	689	728	6?
1851	886	889	823	774	675	786	558	525	481	448	628	576	614	654	604	637	6?
1852	999	937	677	758	774	720	634	747	531	500	476	440	617	559	601	644	5?
1853	1.020	978	880	877	662	739	767	707	628	737	530	498	475	439	613	555	5?
1854	1.059	1.055	920	929	874	869	658	736	764	705	627	736	530	497	474	439	6?
1855	1.075	982	949	986	910	908	871	864	655	732	764	705	626	735	529	496	4?
1856	960	948	989	939	939	976	904	903	867	862	650	728	761	701	623	732	5?
1857	931	969	925	899	978	926	920	969	891	891	849	856	649	722	760	698	6?

Source: http://gamli.hagstofa.is/pages/2947?src=/dm/pyramid/man.html
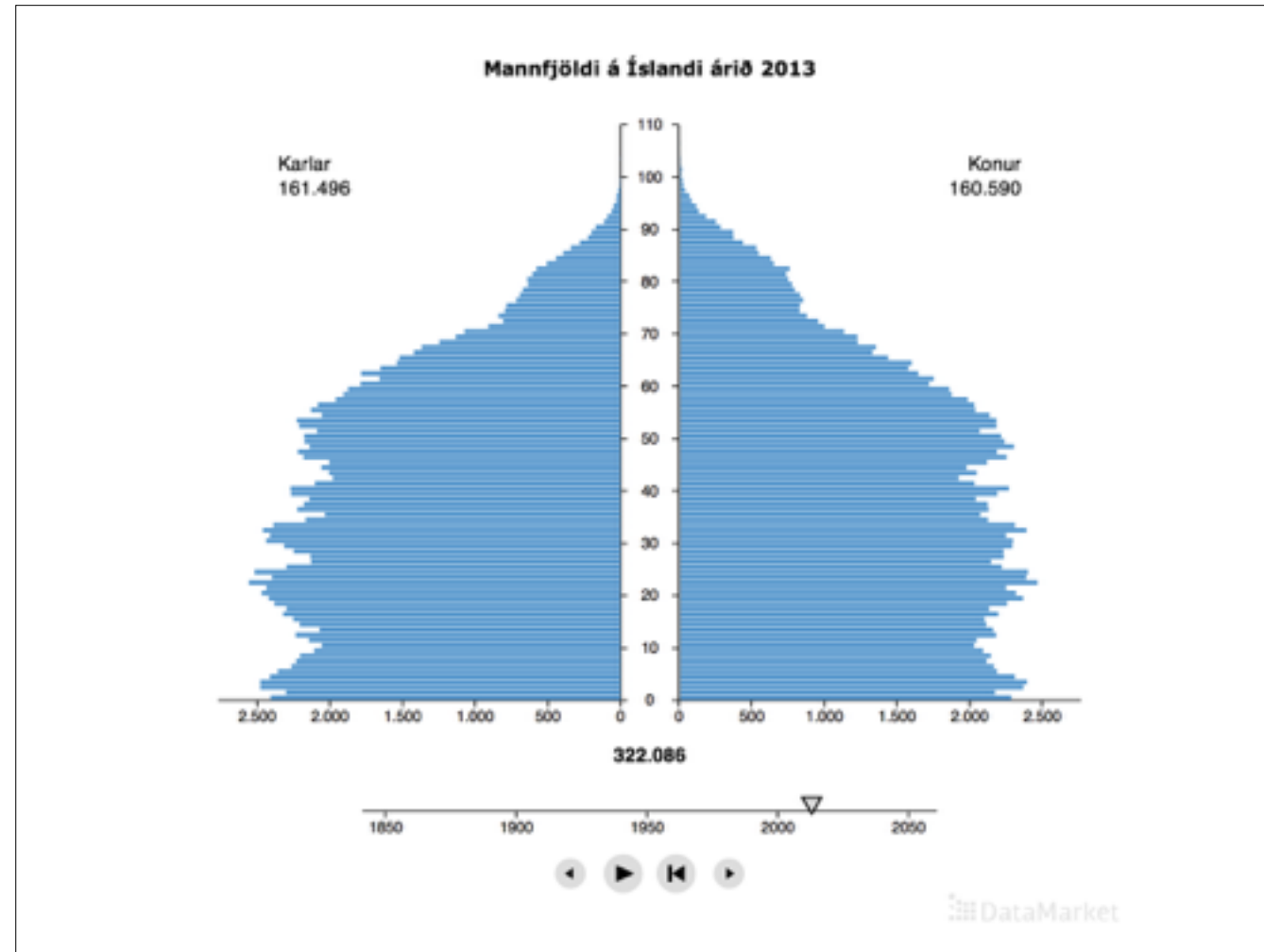
"Laun fullvinnandi launamanna á almennum vinnumarkaði eftir starfi og "
"kyni 1998-2012   "

" "," "," ","Meðaltal"," "
" "," "," ","1 Störf stjórnenda","121 Forstjórar og aðalframkvæmdastjórar fyrirtækja og stofnana","122 Yfirme
"Heildarlaun"
" ","Alls"
" "," ","2000",372,535,322,338,340,359,..,354,362,441,..,..,353,..,..,312,317,..,..,..,318,..,..,301,375,307,337
" "," ","2001",402,556,352,368,369,390,..,392,397,473,..,..,402,..,..,335,344,..,..,..,346,..,..,327,414,325,342
" "," ","2002",432,603,378,391,378,425,..,392,434,501,415,..,424,..,..,315,361,..,..,..,359,..,291,336,427,347,3
" "," ","2003",460,662,397,416,398,440,..,400,448,521,395,..,439,..,..,342,388,..,..,..,384,..,315,357,460,369,3
" "," ","2004",495,697,431,467,425,474,..,413,488,548,438,..,482,..,..,372,413,..,..,..,390,..,326,384,498,390,3
" "," ","2005",614,993,520,529,485,513,..,484,529,579,474,..,527,441,..,472,504,..,..,439,510,587,355,407,527,44
" "," ","2006",720,1151,605,607,556,568,598,532,592,609,502,..,630,442,..,561,589,..,514,490,601,648,377,435,566
" "," ","2007",826,1351,689,668,625,616,644,589,641,682,573,570,660,483,..,644,667,684,536,519,709,835,401,475,6
" "," ","2008",902,1618,710,719,684,676,726,631,717,776,627,641,739,523,..,703,722,762,645,611,750,857,434,499,6
" "," ","2009",749,1121,631,690,604,631,659,606,653,719,595,696,634,524,..,598,604,847,592,553,605,691,439,463,6
" "," ","2010",772,1164,659,708,619,643,681,624,657,731,636,646,639,519,659,612,614,759,593,583,616,736,466,478,
" "," ","2011",843,1293,715,742,666,683,784,665,692,772,674,696,670,545,698,665,673,930,670,656,667,749,482,517,
" "," ","2012",904,1408,730,804,716,739,881,726,740,849,831,696,712,578,736,709,747,912,677,722,747,911,518,564,

Source: gamli.hagstofa.is/Pages/2948/?src=/dm/bars/laun.html

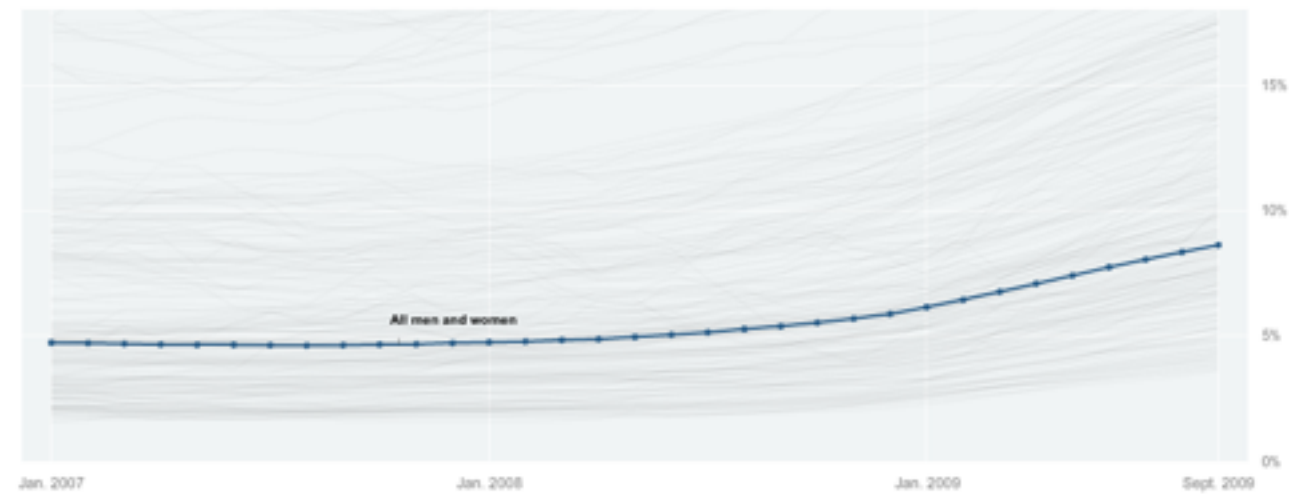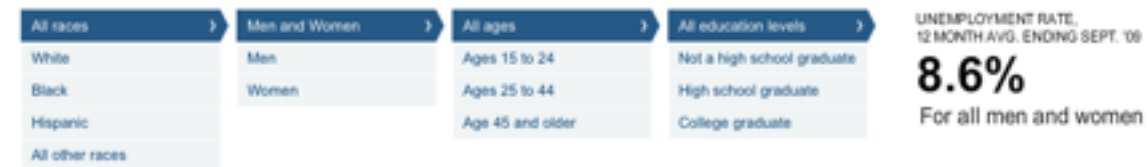| Nr	Dags.	Timi	Breidd	Lengd	Dypi	M	ML
1	20100329	004703.904	63.63744	-19.52769	3.515	1.63	1.24
2	20100329	005117.048	64.43972	-17.70476	6.647	1.66	1.43
3	20100329	005751.922	64.71282	-17.33068	9.363	2.10	2.20
4	20100329	012305.934	64.71010	-17.33636	20.000	0.97	1.12
5	20100329	014818.766	63.63352	-19.52757	4.573	1.19	0.90
6	20100329	014918.823	63.63732	-19.53381	2.800	0.71	0.61
7	20100329	020640.132	63.93967	-22.01205	9.823	0.72	0.43
8	20100329	020719.777	63.63134	-19.55590	1.145	1.10	0.92
9	20100329	022019.963	63.63221	-19.53412	1.145	0.63	0.42
10	20100329	023614.270	64.35008	-20.53393	9.577	0.66	0.69
11	20100329	024046.810	64.34958	-20.52434	9.965	0.77	0.98
12	20100329	025026.434	64.35082	-20.52083	10.990	0.61	0.76
13	20100329	031524.069	63.63536	-19.52435	0.176	0.53	0.52
14	20100329	034622.611	65.21228	-16.27288	3.867	0.72	0.68
15	20100329	035047.982	63.63678	-19.51575	4.045	1.01	0.90
16	20100329	035617.268	63.63778	-19.52523	5.864	1.92	1.37
17	20100329	035708.612	63.63490	-19.52247	4.082	1.02	0.67
18	20100329	040210.033	63.96873	-20.87933	3.429	0.28	0.07
19	20100329	040343.867	63.63396	-19.51731	4.540	1.07	0.94
20	20100329	040356.063	63.63494	-19.51859	3.928	1.78	1.22
21	20100329	040418.359	63.63647	-19.53803	1.589	1.23	0.70
22	20100329	040524.254	63.63660	-19.53869	3.802	1.69	1.03
23	20100329	040757.106	63.96908	-20.87963	3.743	0.88	0.56
24	20100329	042153.765	64.35866	-20.52919	11.001	0.66	0.71
25	20100329	055346.423	63.62316	-19.58168	9.067	1.16	1.02
26	20100329	060915.423	63.96679	-20.88493	3.984	0.24	-0.05
27	20100329	063839.912	64.35296	-20.52490	9.591	0.64	0.88
28	20100329	075531.847	64.35291	-20.51740	11.420	0.68	0.82
29	20100329	075811.699	64.70893	-17.33532	8.382	1.43	1.56
30	20100329	075851.815	64.34653	-20.53815	10.566	0.81	0.87
31	20100329	080202.732	64.35088	-20.53779	10.109	1.19	1.19
32	20100329	080307.062	63.63304	-19.50804	1.379	0.43	0.08
33	20100329	090512.323	64.35028	-20.51835	11.677	0.69	0.80
34	20100329	092453.147	64.35399	-20.52991	9.966	0.91	1.08
35	20100329	094734.730	63.96495	-20.88145	4.211	0.34	0.30
36	20100329	102823.149	63.63066	-19.52735	1.603	0.90	1.96
37	20100329	102829.289	63.63709	-19.53898	3.819	1.82	1.74
38	20100329	103008.710	63.63412	-19.53262	1.286	1.07	1.02
39	20100329	103031.923	63.63321	-19.53086	2.491	1.02	0.86
40	20100329	103143.608	63.63271	-19.50310	9.827	1.04	0.48
41	20100329	104128.568	66.25287	-16.74530	5.009	0.73	1.03
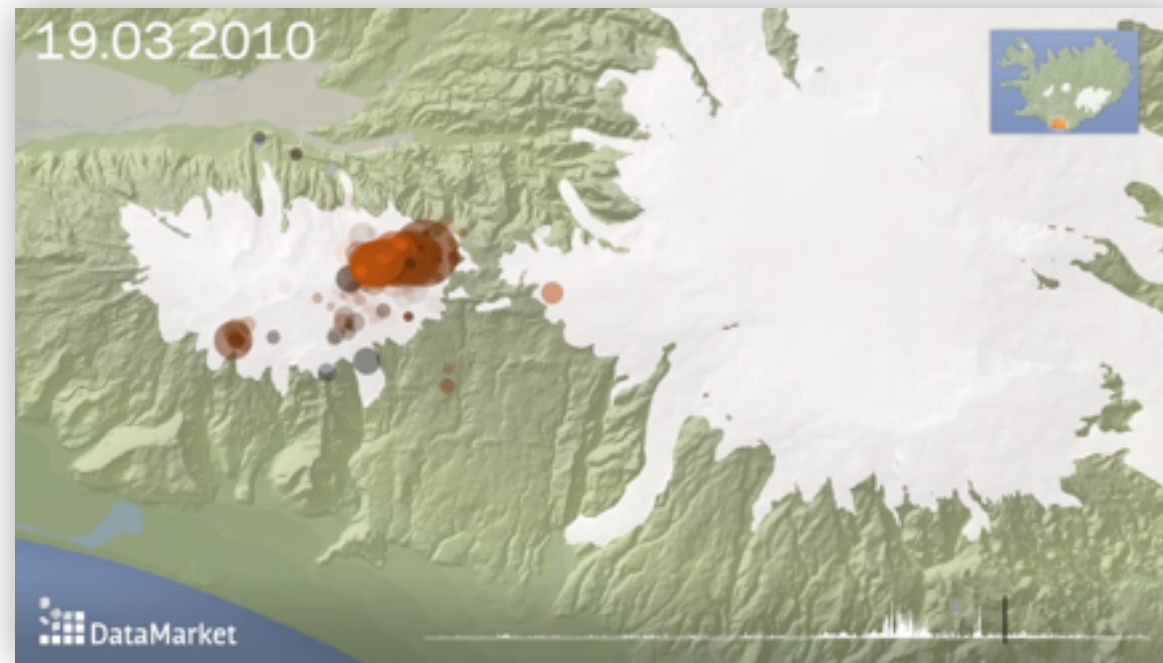
Source: datamarket.com/is/gallery/earthquakes_and_eruptions/

More examples

*„If we have data, let's look at data. If all we
have are opinions, let's go with mine."*
— Jim Barksdale

The Jobless Rate for People Like You

Not all groups have felt the recession equally.

All races	Men and Women	All ages	All education levels
White	Men	Ages 15 to 24	Not a high school graduate
Black	Women	Ages 25 to 44	High school graduate
Hispanic		Age 45 and older	College graduate
All other races			

UNEMPLOYMENT RATE,
12 MONTH AVG. ENDING SEPT. '09

8.6%
For all men and women

All men and women

15%

10%

5%

0%

Jan. 2007   Jan. 2008   Jan. 2009   Sept. 2009

Source: www.nytimes.com/interactive/2009/11/06/business/economy/unemployment-lines.html

Budget Forecasts, Compared With Reality

Just two years ago, surpluses were predicted by 2012. How accurate have past White House budget forecasts been?

1 2 3 4 5 6 NEXT ▶

Falling short

President Obama's budget proposal estimates a deficit of $1.6 trillion for the current fiscal year and $1.3 trillion in 2011.

↑ Annual surplus   In today's dollars
↓ Or deficit

$0.0 trillion

-0.5

-1.0

2011 proposal: -$1.27

-1.5

2010 estimate: -$1.56

1982   1986   1990   1994   1998   2002   2006   2010   2020
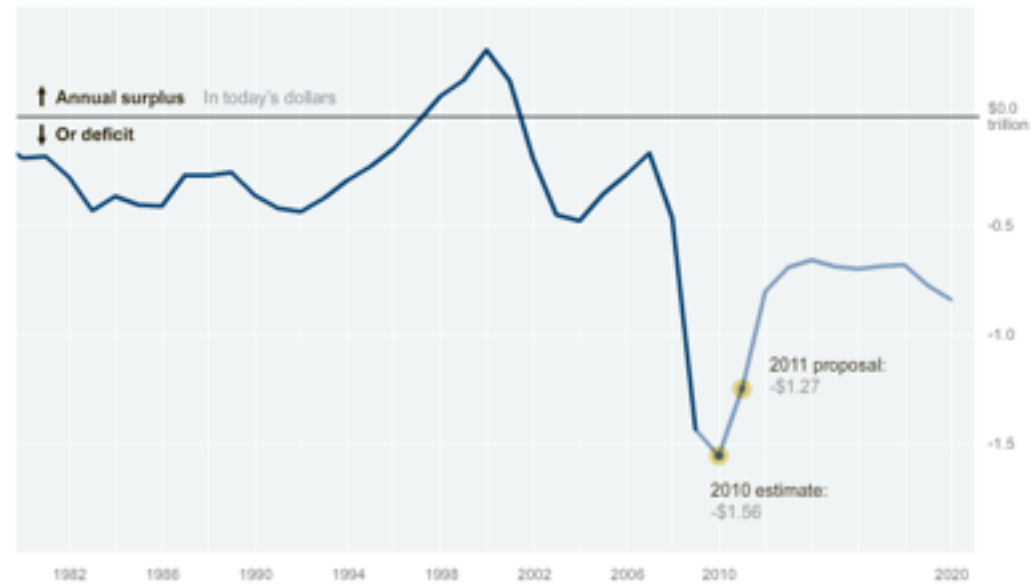
Source: www.nytimes.com/interactive/2010/02/02/us/politics/20100201-budget-porcupine-graphic.html

Source: www.youtube.com/watch?v=z50XoqzOBLQ
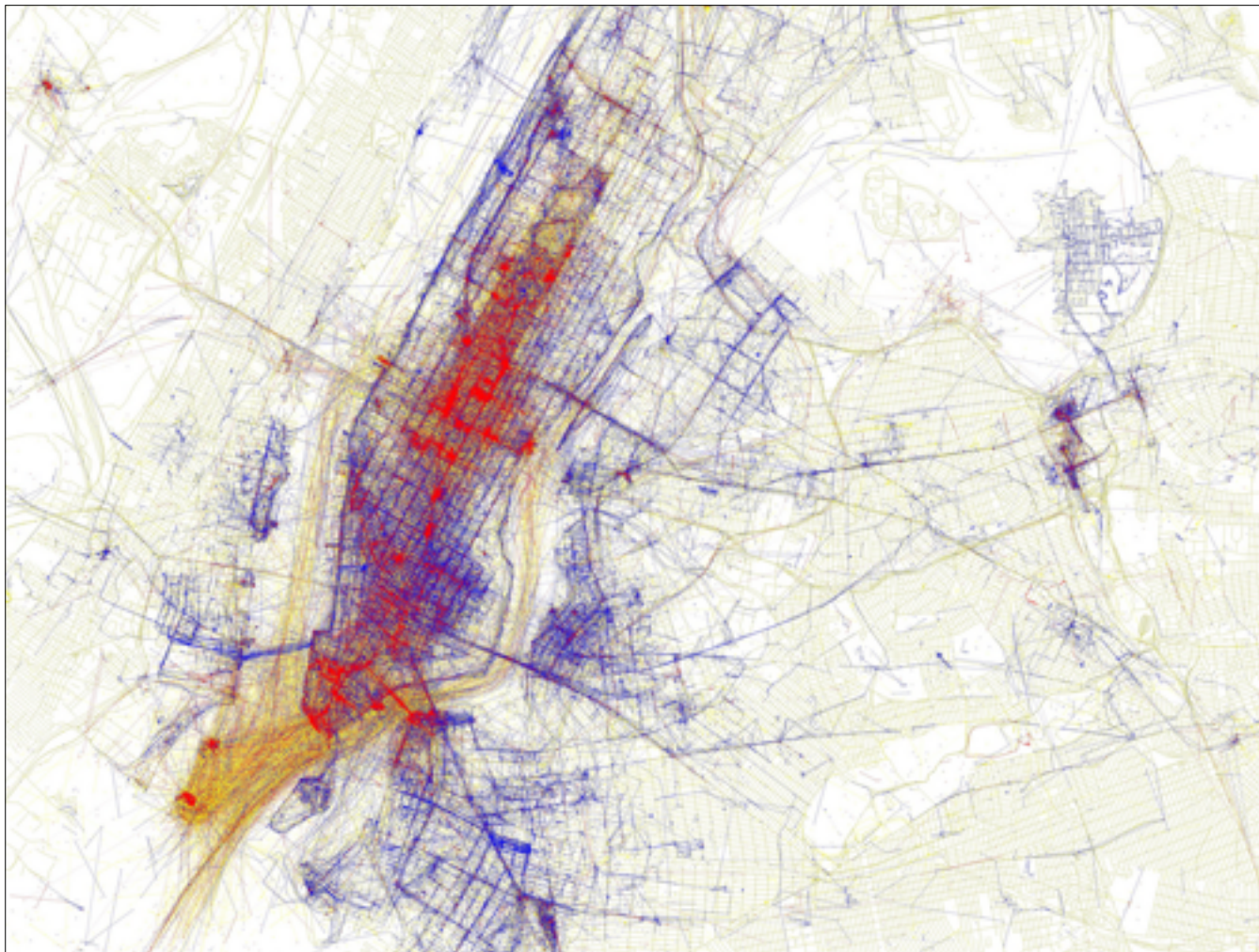
Source: https://www.flickr.com/photos/walkingsf/4671594023/in/set-72157624209158632

Iraq's bloody toll

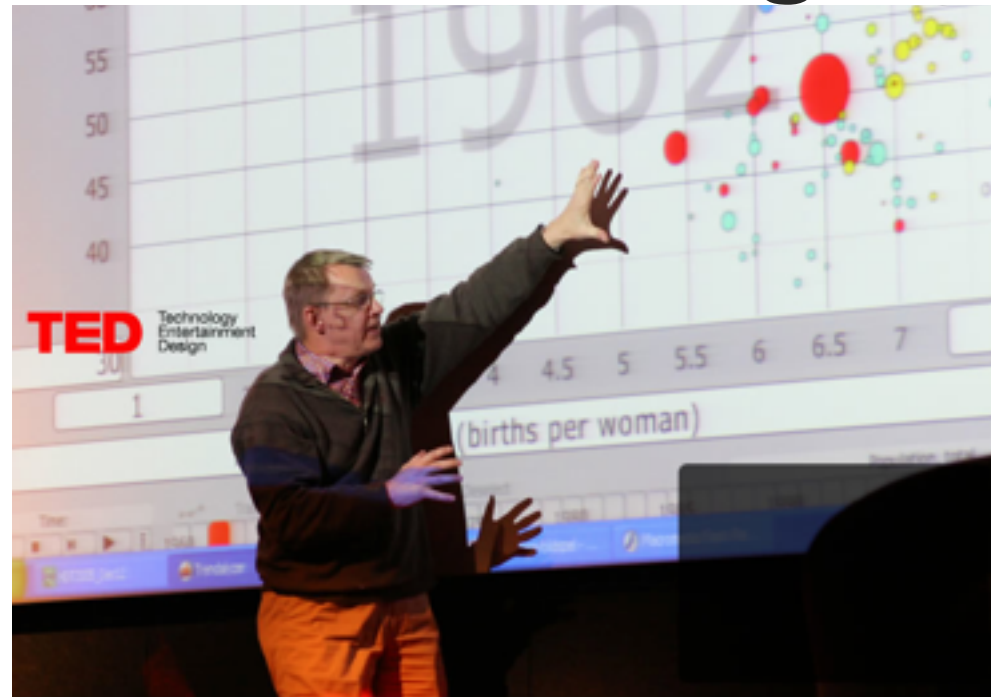Source: http://www.scmp.com/infographics/article/1284683/iraqs-bloody-toll
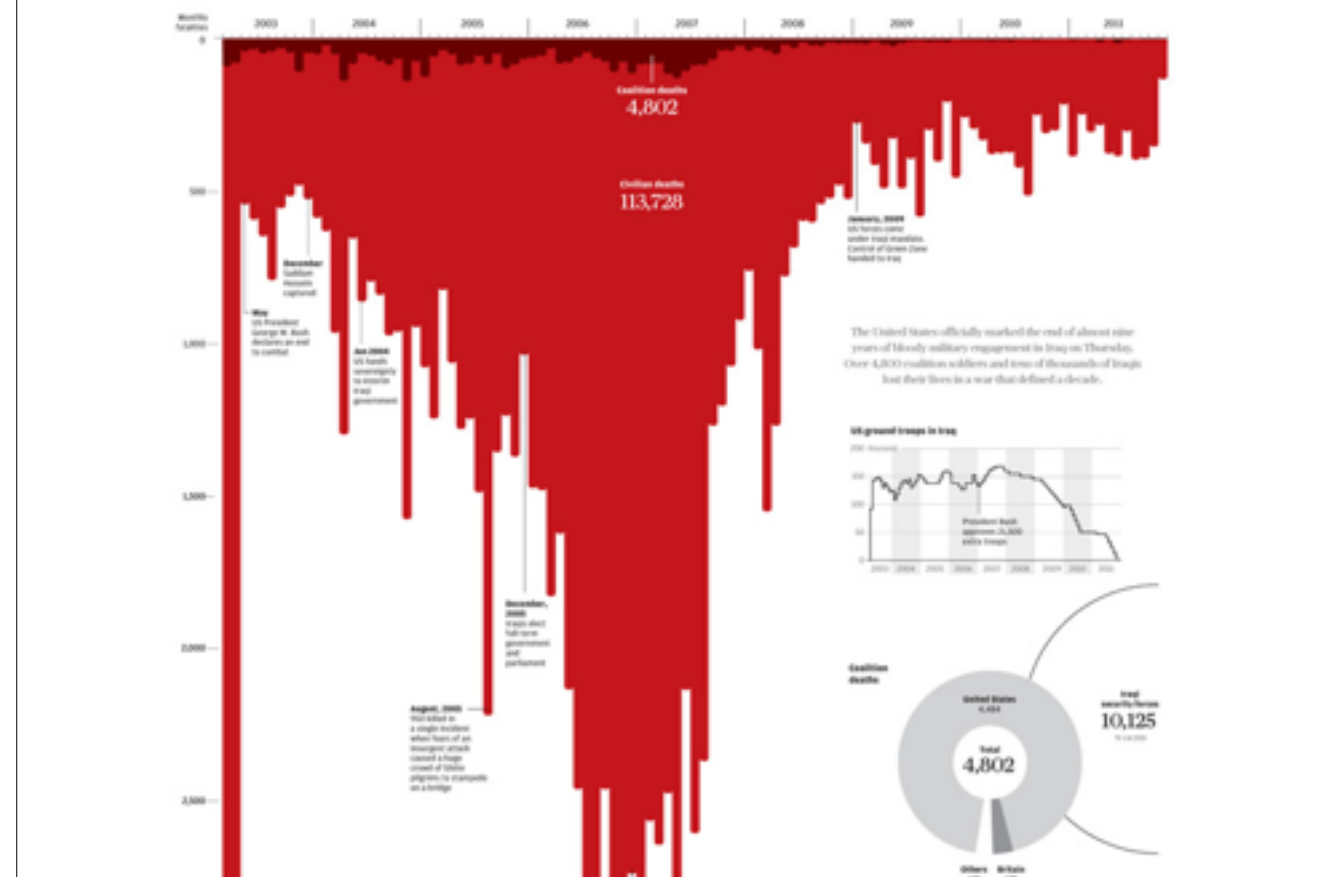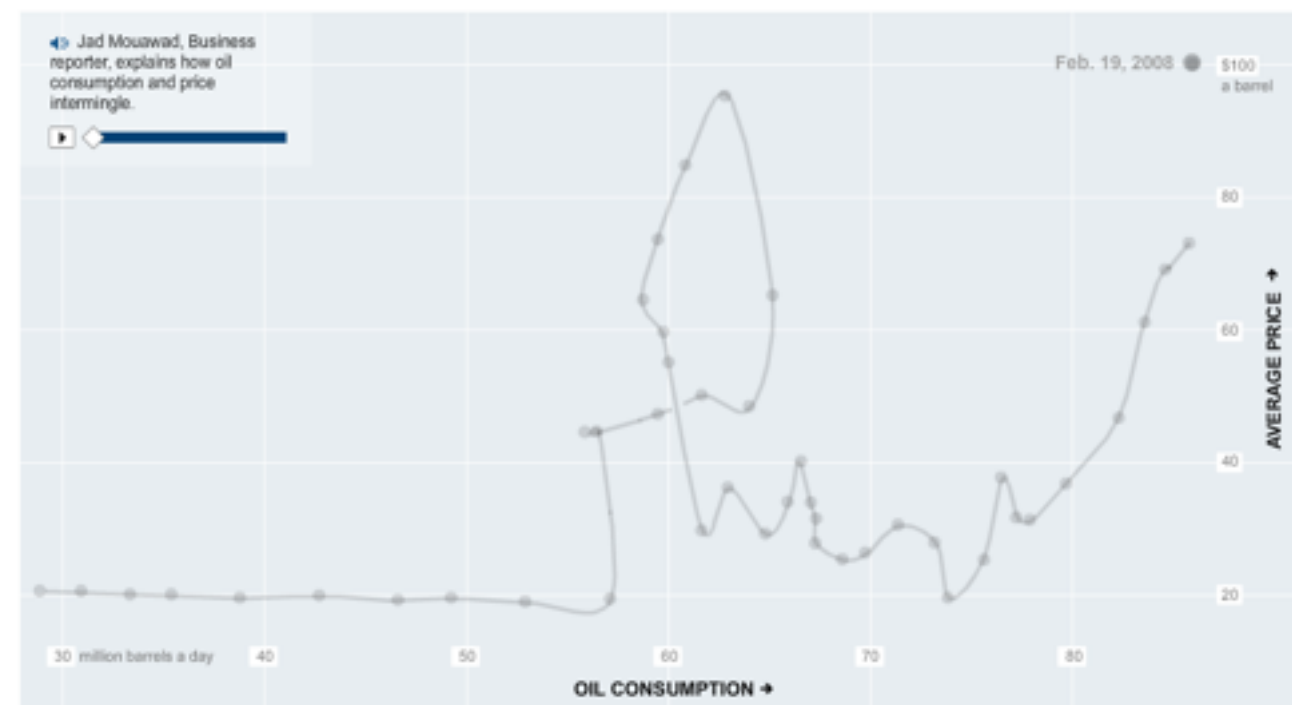
Source: www.nytimes.com/interactive/2008/02/19/business/20080220_CENTURY_GRAPHIC.html
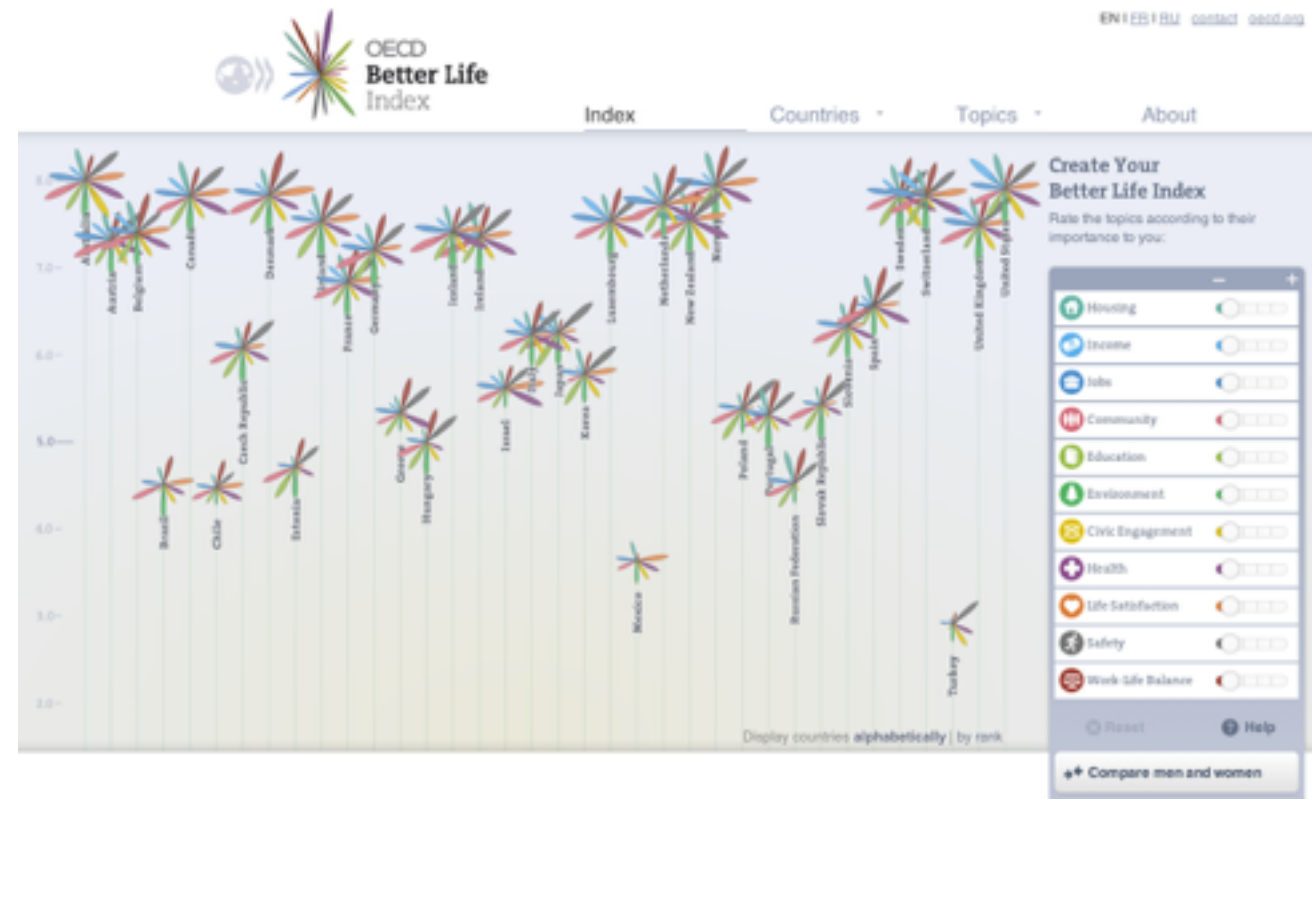
Source: www.oecdbetterlifeindex.org/

Hans Rosling

Source: www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html

Source: www.gapminder.org/world/

Tools and Inspiration

„I could give you facts and figures.
Even give you plans and forecasts“

— Jesus in Jesus Christ Superstar

Tools: Software

- Regular expressions / grep
- Advanced text editors: TextMate, TextWrangler
- Excel (yes, Excel!)
- Open Refine
- R
- DataGraph
- Tableau
- Illustrator (vector graphics)
- Photoshop (pixel graphics)

Protovis, Processing, amCharts, HighCharts, Flot, Raphael, Flash, Google Maps API

Tools: Finding data

- Data.gov.uk/data
- Data.gov/catalog
- DataMarket/data/
- US Census

- OpenStreetMap
- GeoNames

DataMarket, Data.gov.uk, Data.gov

Tools: Programming

- Processing and Processing.js
- amCharts, HighCharts, Flot
    - A list of 25 chart components
    - A list of 15 chart components
- Google Chart Tools
- Google Maps API
- Raphaël and gRaphaël
- Protovis
- D3

Blogs

- [Flowing Data](#)
- [Datavisualization.ch](#)
- [Well formed data](#)
- [Information Aesthetics](#)
- [Perceptual Edge](#)
- [The Why Axis](#)
- [Fell in love with data](#)

- [Junk Charts](#)

FlowingData, Datavisualization.ch, Information Aesthetics, Junk Charts

Books

- **The Functional Art**, Alberto Cairo
- **Visualizing Data**, Ben Fry
- **Show Me the Numbers**, Stephen Few
- **Now You See It**, Stephen Few
- **The Visual Display of Quantitative Information**, Edward Tufte
- **The WSJ Guide to Information Graphics**, Dona M. Wong
- **Visualize this**, Nathan Yau

- **Don't Make Me Think**, Steve Krug
- **Presentation Zen**, Garr Reynolds