

Data visualization and storytelling

Jens Finnäs
Journalism++ Stockholm



14/12

Introduction + Excel and Infogram exercises.

17/12

Workshop about finding and researching the story in data with **Paloma Pérez Lucero** from TT.

18/1

Visualizing lots of data on small screens with **Henrik Hedvall** from Tink.

20/1

Workshop about visualizing data with **Information is beautiful**.

22/1

Mid-term presentations

4/2

Final presentations



Clients and briefs

Inizio/Aftonbladet: How can we understand the raise of the Sweden democrats?

Nacka kommun: How to counter gender inequality with data?

Helsingborg stad: Why are we growing? And how does it affect our city?



The secret to good data visualization

...and why we have this course.

@jensfinnas

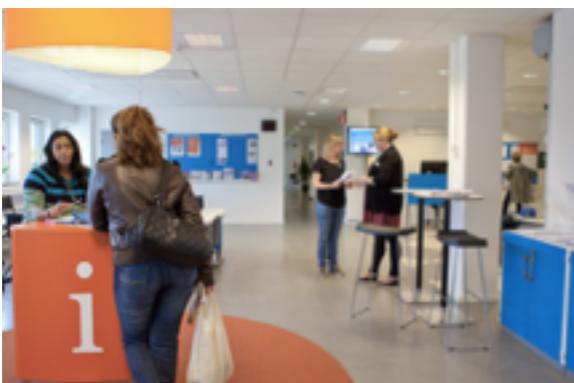


To understand society you have
to understand data.

@jensfinnas



Society → Data



	A	B	C
1	Arbetslöshet		
2	Andel av arbetskraften, 15-74 år		
3	Månad	Procent	Procent, säsong
4	jan-01	6,4	5,8
5	feb-01	6,1	5,8
6	mar-01	6,1	5,8
7	apr-01	5,8	5,8
8	maj-01	6,0	5,8
9	jun-01	6,8	5,8
10	jul-01	5,5	5,8
11	aug-01	5,4	5,8
12	sep-01	5,4	5,9
13	okt-01	5,6	5,9
14	nov-01	5,3	5,9
15	dec-01	5,5	5,9
16	jan-02	6,2	5,9
17	feb-02	6,1	5,9
18	mar-02	6,2	5,8

@jensfinnas



To understand data you have to
be able to **visualize data**.

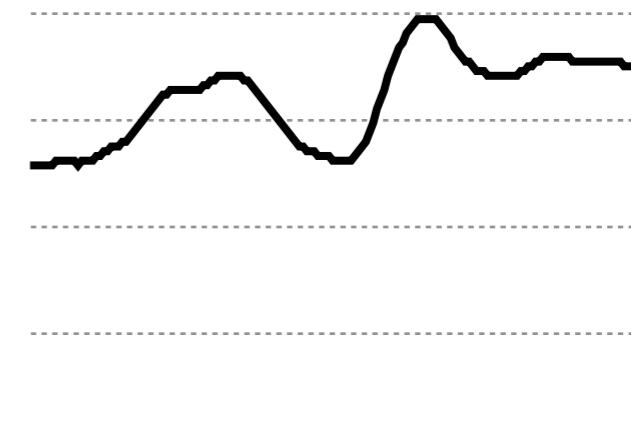
@jensfinnas



Society → Data → Visualization



	A	B	C
1	Arbetslöshet		
2	Andel av arbetskraften, 15-74 år		
3	Månad	Procent	Procent, säsong
4	jan-01	6,4	5,8
5	feb-01	6,1	5,8
6	mar-01	6,1	5,8
7	apr-01	5,8	5,8
8	maj-01	6,0	5,8
9	jun-01	6,8	5,8
10	jul-01	5,5	5,8
11	aug-01	5,4	5,8
12	sep-01	5,4	5,9
13	okt-01	5,6	5,9
14	nov-01	5,3	5,9
15	dec-01	5,5	5,9
16	jan-02	6,2	5,9
17	feb-02	6,1	5,9
18	mar-02	6,2	5,8





So what is the secret
good data visualization?

@jensfinnas



```
require 'rubygems'
require 'nokogiri'
require 'open-uri'

friends = Array.new

list = Array.new
list = ["test"]

def do_stuff1()

list_of_blogs = Nokogiri::HTML(open("http://www.ratata.fi/alla_bloggar/#front-page?offset_3472=1230"))

fetch_links = list_of_blogs.xpath("//div[@class='col w50 title']/a").collect[1..-1]

fetch_links.each do |link_to_test|
  full_url = link_to_test["href"]
  list.push(full_url)
  puts full_url
end

list.each_index do |i|
  id = i
  link = list[i]
  File.open("ratata.txt", 'a'){ |f|
    f.write("#{id}\t#{link}\n")
  }
end
end
```

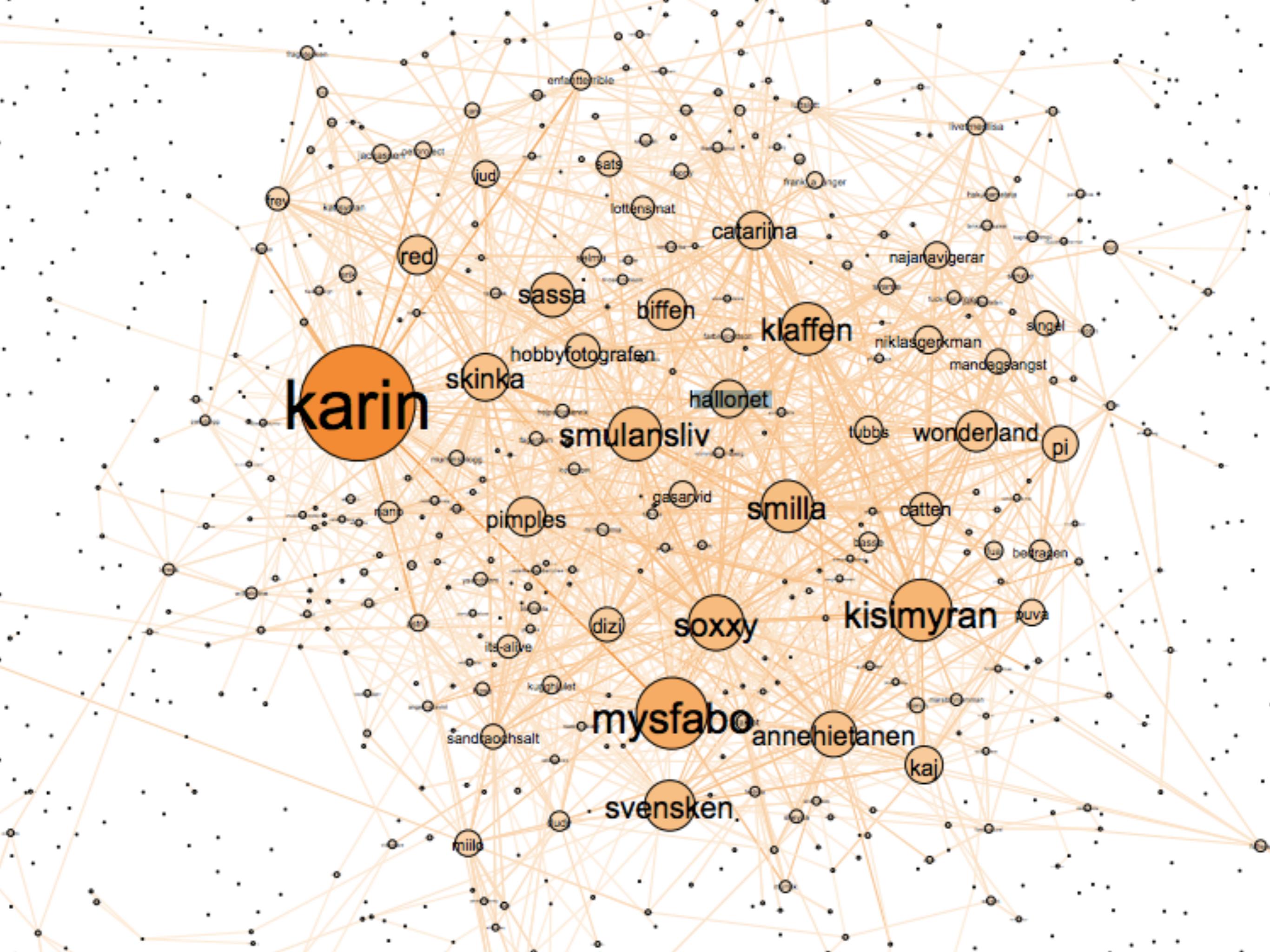
@jensfinnas



A	B	C	D
1	0	http://newstone.ratata.fi	
2	1	http://miruschkan.ratata.fi	
3	2	http://ysandstrm.ratata.fi	
4	3	http://henriksdotter.ratata.fi	
5	4	http://mickeystar.ratata.fi	
6	5	http://peoples.ratata.fi	
7	6	http://villarosa.ratata.fi	
8	7	http://lindas.ratata.fi	
9	8	http://pamsus.ratata.fi	
10	9	http://thatsallfolks.ratata.fi	
11	10	http://puusan.ratata.fi	
12	11	http://hsb.ratata.fi	
13	12	http://johannah.ratata.fi	
14	13	http://matiaslofgren.ratata.fi	
15	14	http://janey.ratata.fi	
16	15	http://chanettes.ratata.fi	
17	16	http://sofiasd.ratata.fi	
18	17	http://tt.ratata.fi	
19	18	http://susolo.ratata.fi	
20	19	http://papperslapp.ratata.fi	
21	20	http://teberoende.ratata.fi	
22	21	http://armito.ratata.fi	
23	22	http://formicarufmas.ratata.fi	
24	23	http://paulas.ratata.fi	
25	24	http://visualisten.ratata.fi	
26	25	http://weheartlina.ratata.fi	

A	B
1	1 550
2	1 764
3	1 781
4	2 233
5	2 366
6	2 1205
7	2 435
8	2 1142
9	2 691
10	2 1201
11	2 727
12	6 868
13	6 399
14	6 780
15	6 473
16	6 368
17	6 1129
18	12 1076
19	16 37
20	17 716
21	22 119
22	22 362
23	22 878
24	22 294
25	22 569
26	22 1003

@jensfinnas





dataist

a blog about data exploration

← [Political Misuse of Statistics Busted](#)

[Project One: Visualizing Friendship](#) →

Mapping Ratata: Who's Hot?

Posted: February 5, 2011 | Author: [Jens Finnäs](#) | Filed under: [Own projects](#) | Tags: [gephi](#), [network](#), [ratata](#), [ruby](#), [screen scraping](#), [social network](#) | Modify: [Edit](#) | 8 Comments

I wanted to play around in [Gephi](#) a bit more after my previous post about [visualizing my social network on Facebook](#). So for my second project I turned my eyes to [Ratata](#), a Swedish blog community in Finland with just over 1200 bloggers. A friend of mine, [Poppe](#) (also on Ratata), has been talking about analyzing the Swedish blogosphere. I hope he doesn't mind me "borrowing" the idea.

I have almost no prior programming experience, but for some time now I have been trying to learn more about [screen scraping](#). Guided mostly by the [Dan Nguyen's](#) brilliant tutorial on [coding for journalist](#) I have started to know my way around Ruby. [Scraper wiki](#) also provides good guidance for those of us who still mostly do copy-paste programming.

After two days of trial and error I managed to put together a script that extracts all the links to fellow Ratata blogs from all the 1207 blogs. That gave me a data set of almost 2000 connections (due to some technical issues I had to exclude a couple of blogs). I obviously wanted to find out who is

What's this?

This blog is about finding, exploring and presenting data online. Or simply data journalism.

The author is [Jens Finnäs](#), a freelance journalist from Finland, currently living in Stockholm, Sweden.

Get in touch

- jens.finnas@gmail.com
- Twitter: [@jensfinnas](#)
- Twitter: [@the_dataist](#)

Tags

[age](#) [animation](#) [anti-jihadist](#) [athletics](#)
[backlinks](#) [Brevik](#) [campaign](#) [funding](#)
[climate change](#) [copyright laws](#) [crime](#) [d3.js](#)
[earthquake](#) [economy](#) [ehdolla.org](#)

@jensfinnas



What is **bad** data visualization?

@jensfinnas



READING, WRITING, AND EARNING MONEY

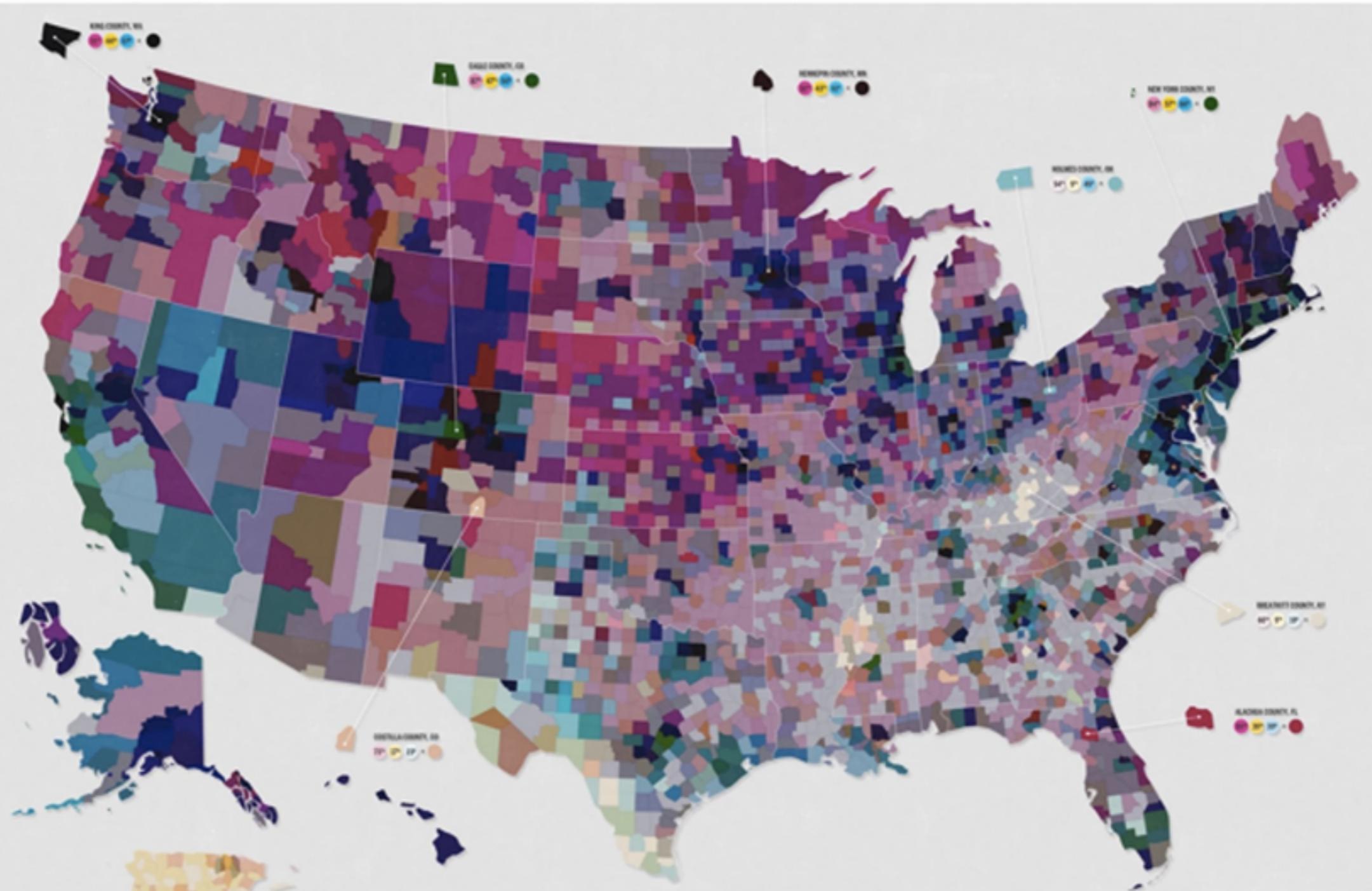
The School Data from the U.S. Census's American Community Survey paints a fascinating picture of the United States at the county level. We've looked at the educational achievement and the median income of the nation's counties, to see where people are going to school, where they're earning money, and if there is any correlation.



The map at right is a product of overlaying the three sets of data. The variations in hue and value has been produced from the data sources above; in general, darker counties represent a more educated, better paid population while lighter areas represent communities with fewer graduates and lower incomes.



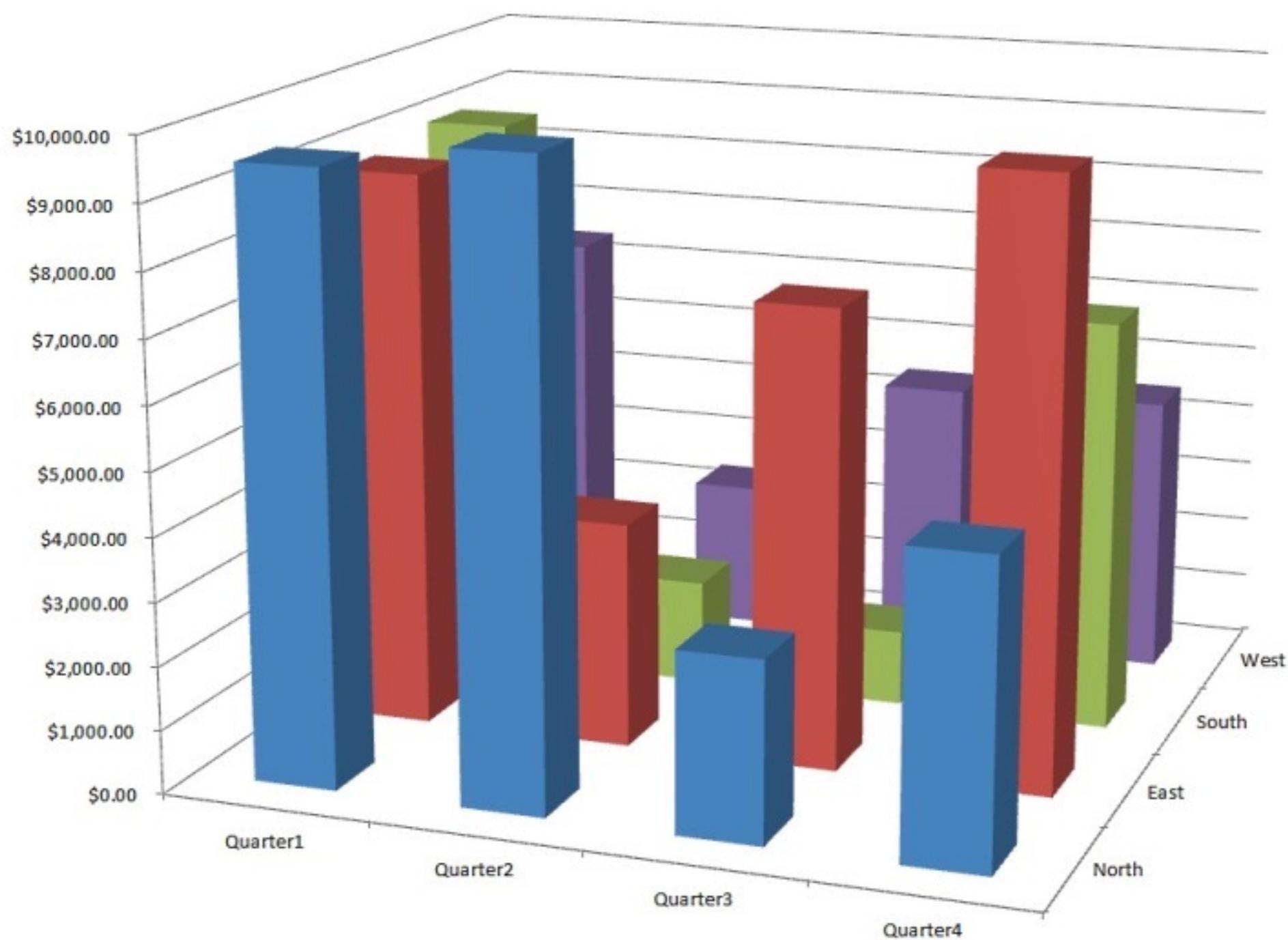
A collaboration between David and Gregory Wolfson.
©2012 E. O. T. Design.



@jensfinnas



Quarterly Report by Region



@jensfinnas



PRIIZE POOLS

2014 DOTA 2
The International



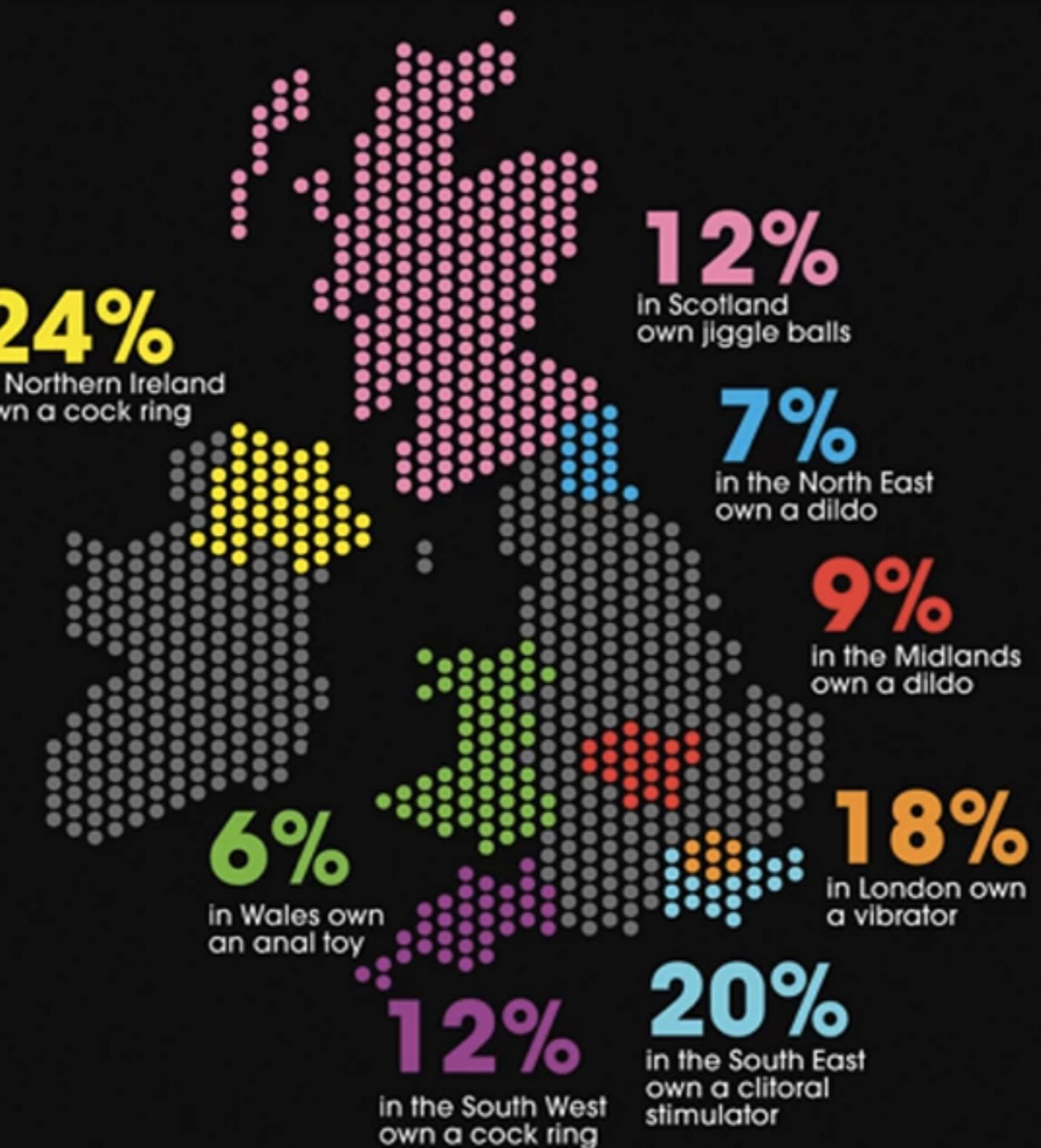
2014
US Masters



Ann Summers

NATIONAL ORGASM DAY

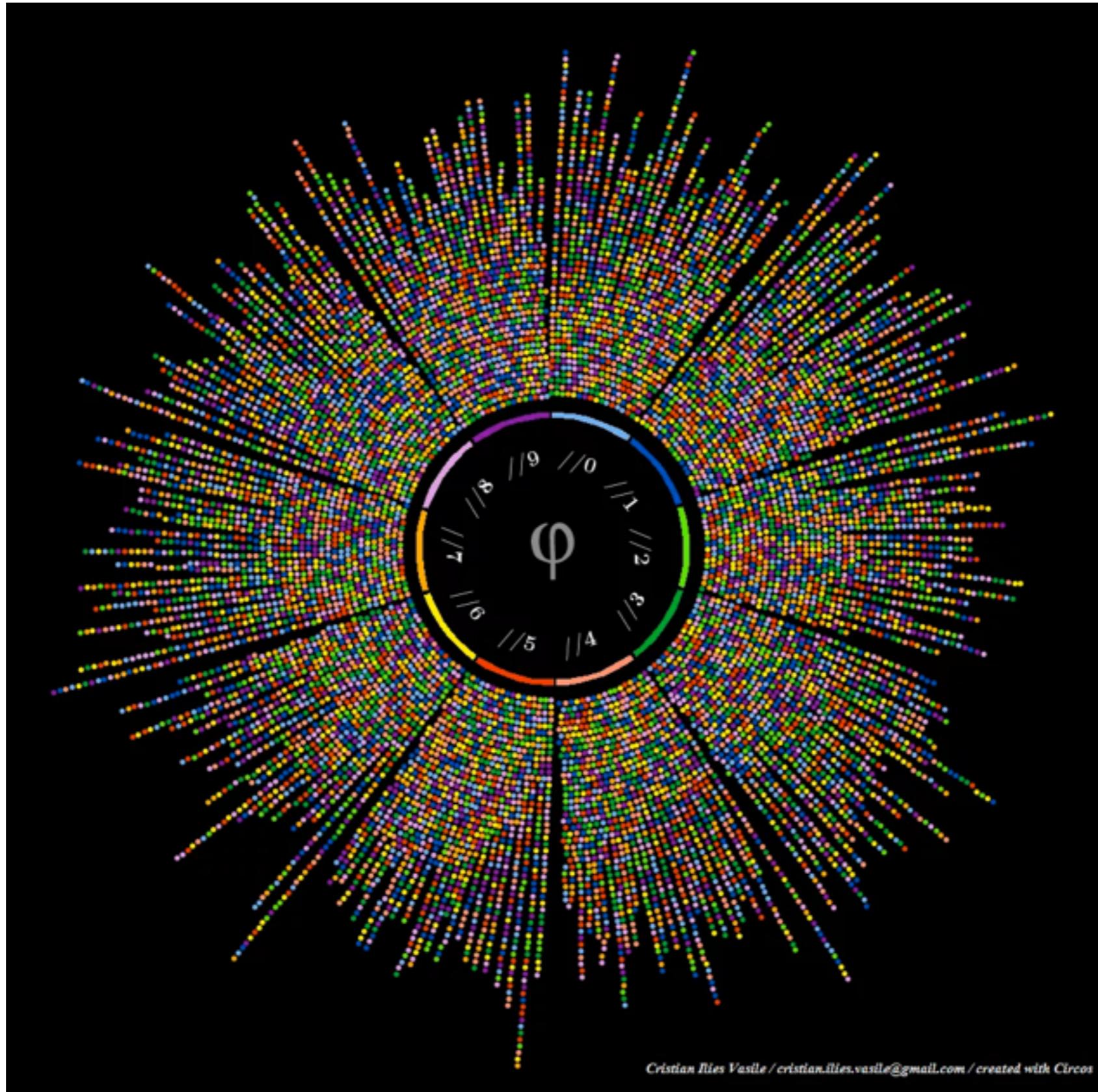
UK SEX TOY HABITAT



@jensfinnas



@jensfinnas



@jensfinnas



@jensfinnas



What makes Hans Rosling's data visualizations great?

@jensfinnas



Digital vs analogue

1. Animation.
2. Interaction.
3. Narration.
4. Boundlessness.



Good data visualization
tells a story.

@jensfinnas



Good data visualization
has something to say.

@jensfinnas



Clients and briefs

Inizio/Aftonbladet: How can we understand the raise of the Sweden democrats?

Nacka kommun: How to counter gender inequality with data?

Helsingborg stad: Why are we growing? And how does it affect our city?



How to do it?

@jensfinnas



Why?



What?



How?



Do!

- **Why** do we do this?
- Who is our **audience**?
- When have we **succeeded**?



Why?



What?



How?



Do!

Is the purpose to...

...explain a phenomenon?

...make an argument?

...shock the user?

...let the user explore?



Why?



What?



How?



Do!

- What **data** do we have?
- What **questions** can we ask our data?



Why?



What?



How?



Do!

Interviewing a dataset requires...

...source criticism.

...an understanding of the topic.

...interview technique.



Why?



What?



How?



Do!

- What **data** do we have?
- What **questions** can we ask our data?
- What is **our story**?



Why?



What?



How?



Do!

Do you want to highlight...

...a **difference**?

...a **change over time**?

...a **relationship**?

...a **geographical pattern**?



- How do we **tell the story**?
- What is the best **format** and **platform**?
- Should (or does it have to) be **interactive**?



Tusen ton koldioxidekvivalenter

25000

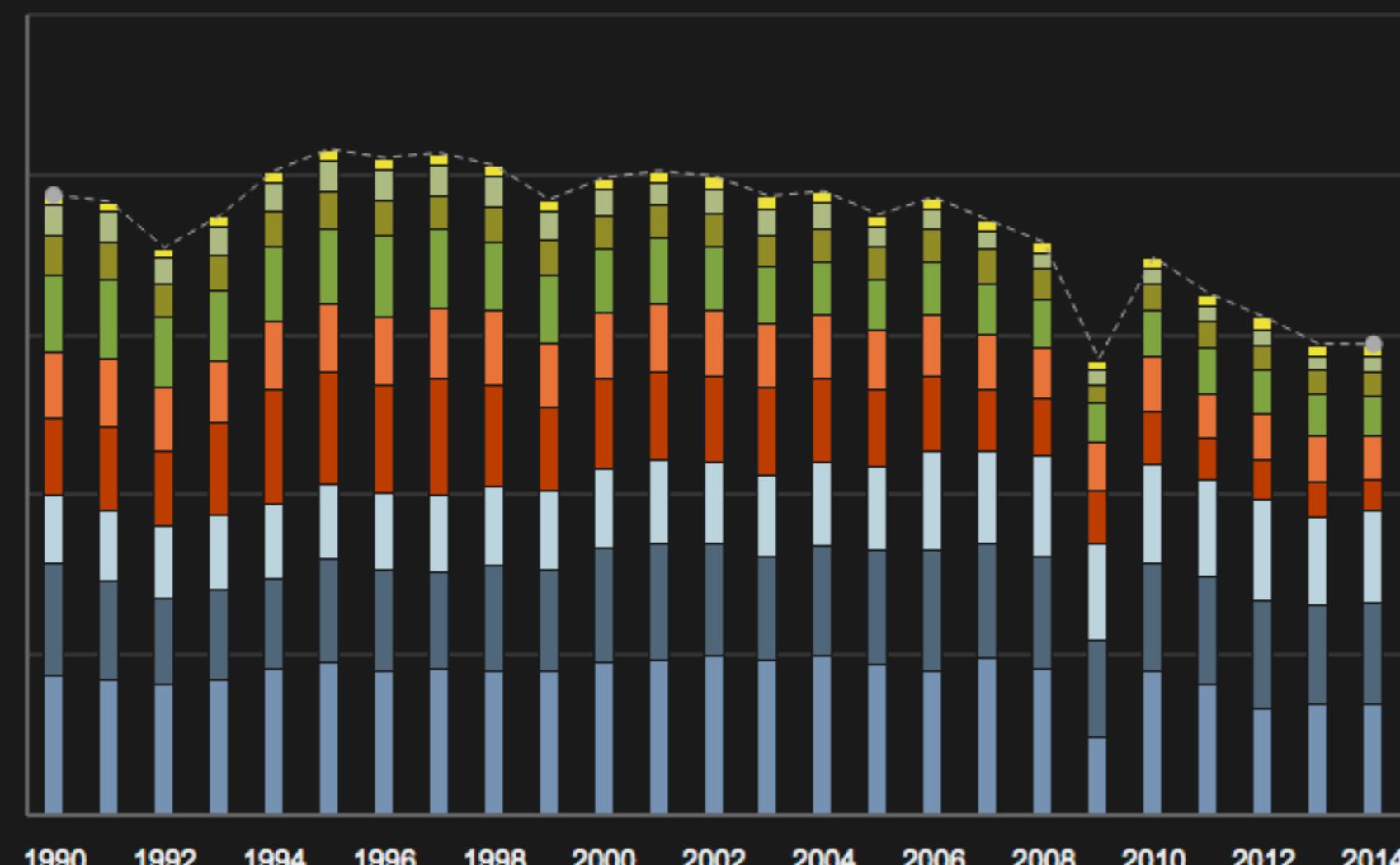
20000

15000

10000

5000

0



Utsläpp av växthusgaser från industrin 1990–2014

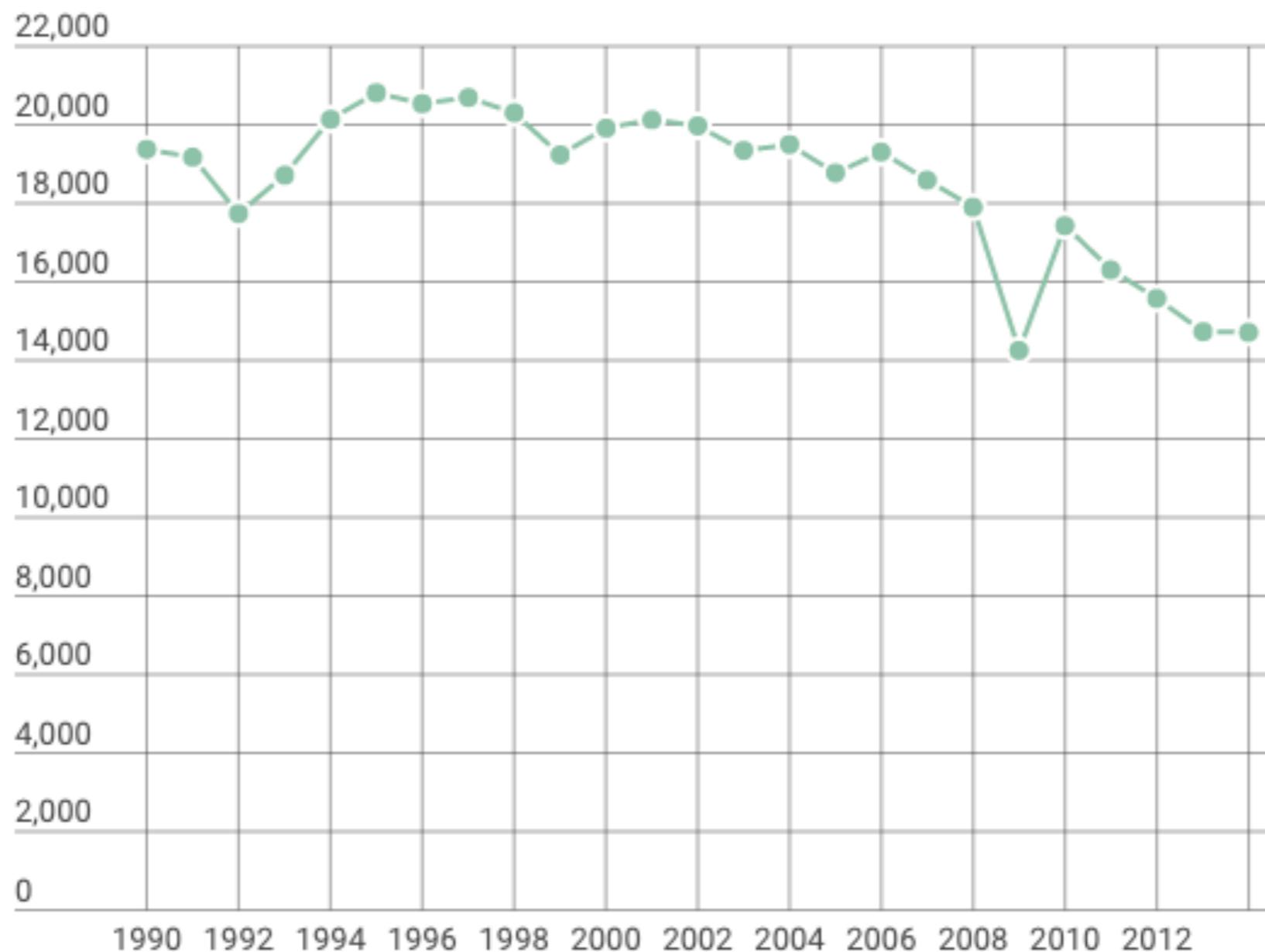
Utsläpp av växthusgaser från Industrin

Tillhör Sveriges officiella statistik.

- Dölj alla
- Koksverk
- Livsmedelsindustri
- Metallindustri (exkl. järn- och stål)
- Övrig (gruvindustri, trävaruhandel m.m.)
- Kemiindustri
- Pappers- och massa-industri samt tryckerier
- Raffinaderier samt distribution av olja och gas
- Mineralindustri (exkl. metaller)
- Järn- och stålindustri
- Totalt

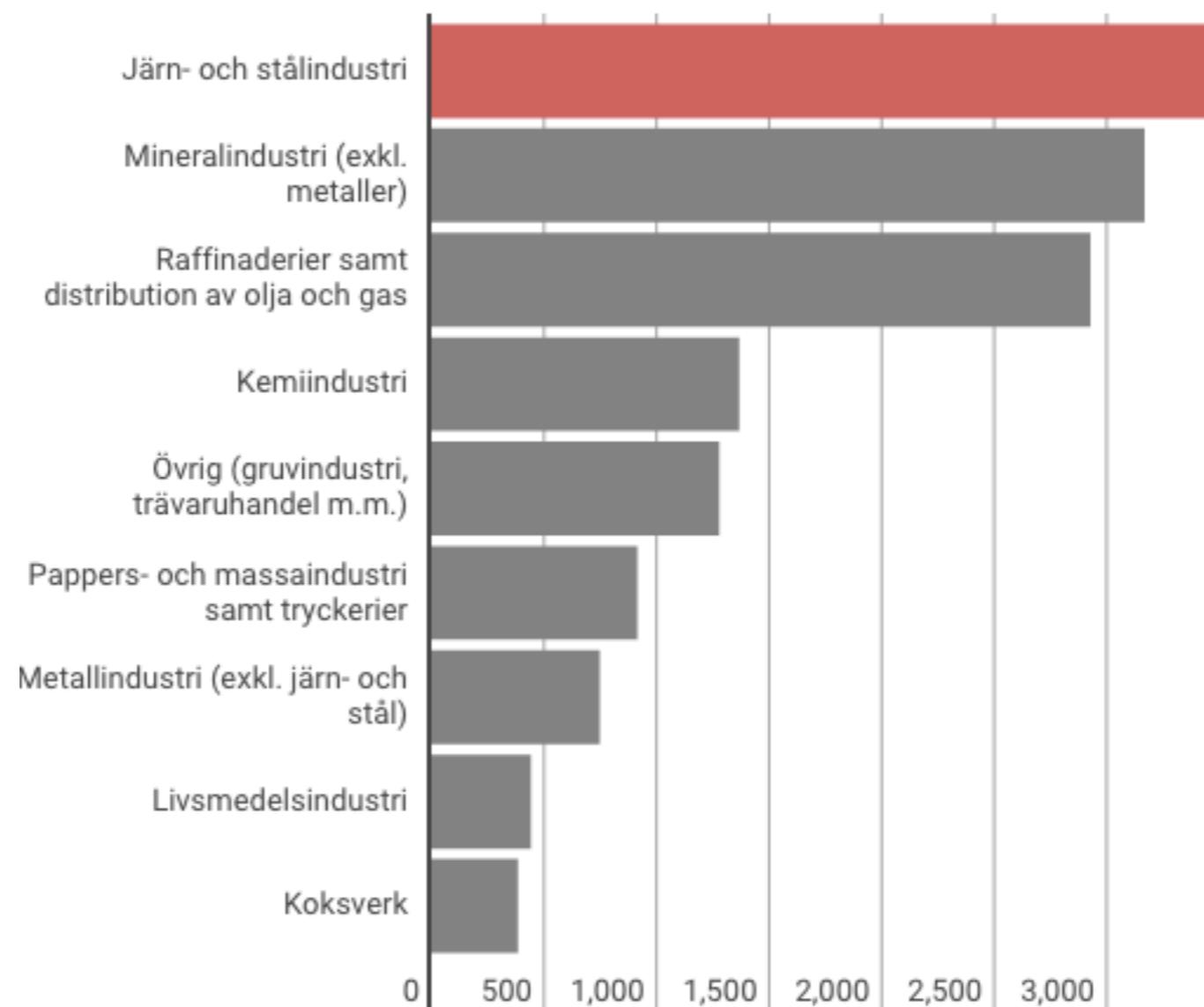


Emissions from industries have decreased by a quarter since 1990



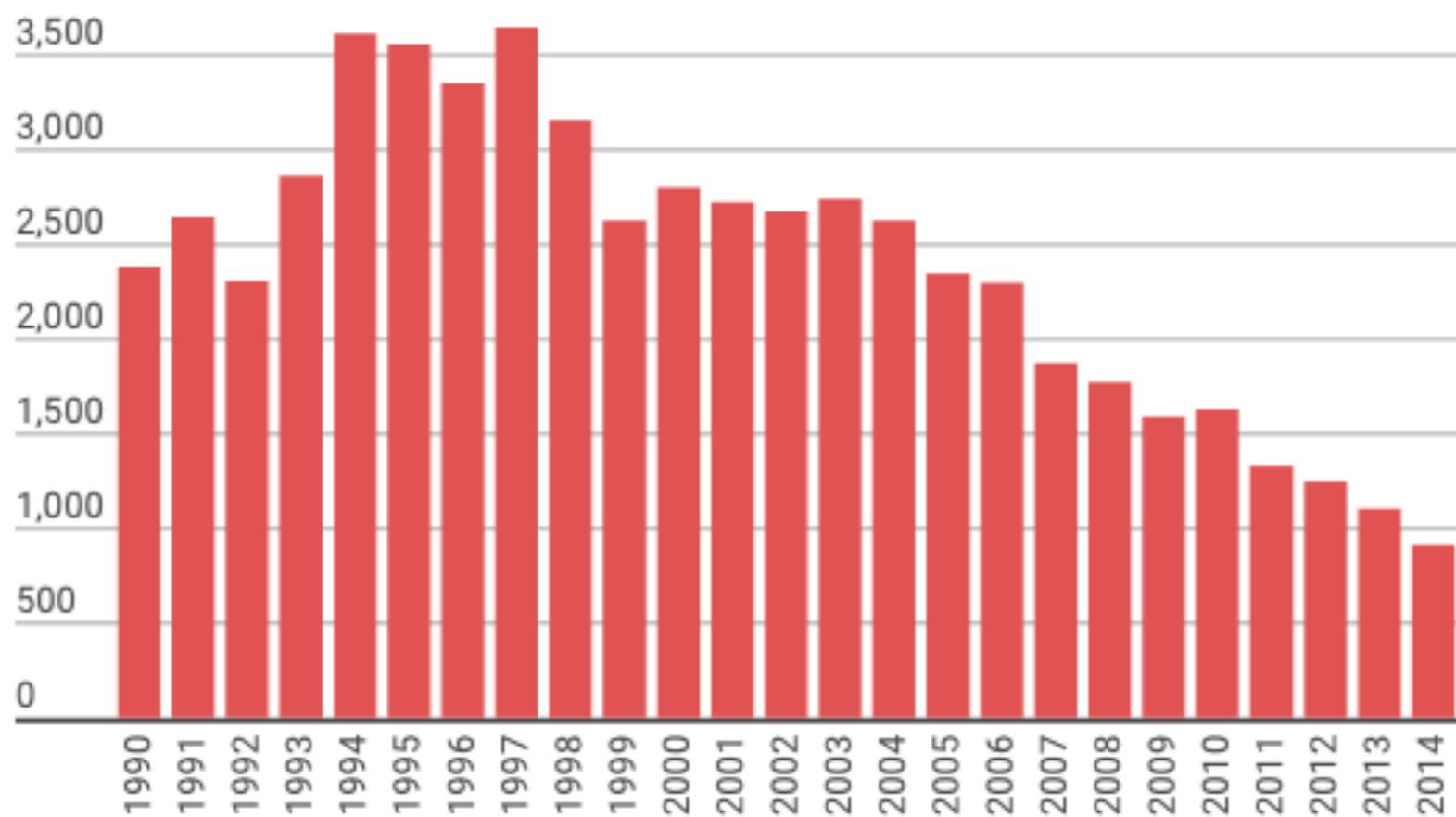


The iron and steel industry is the biggest emitter





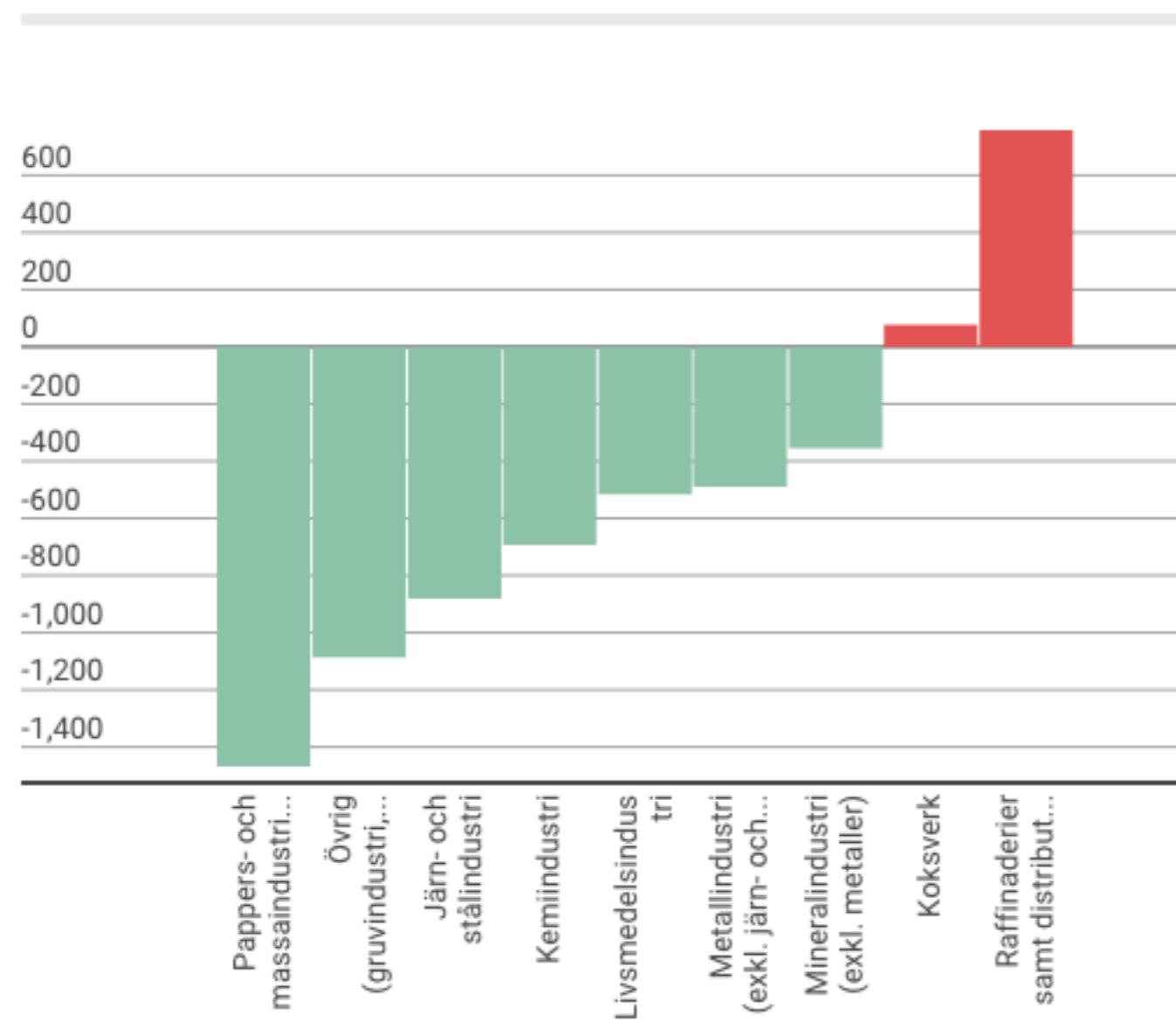
Emissions from the paper industry have decreased heavily since 1990



@jensfinnas



Emissions have decreased in seven of nine industries since 1990





Tusen ton koldioxidekvivalenter

25000

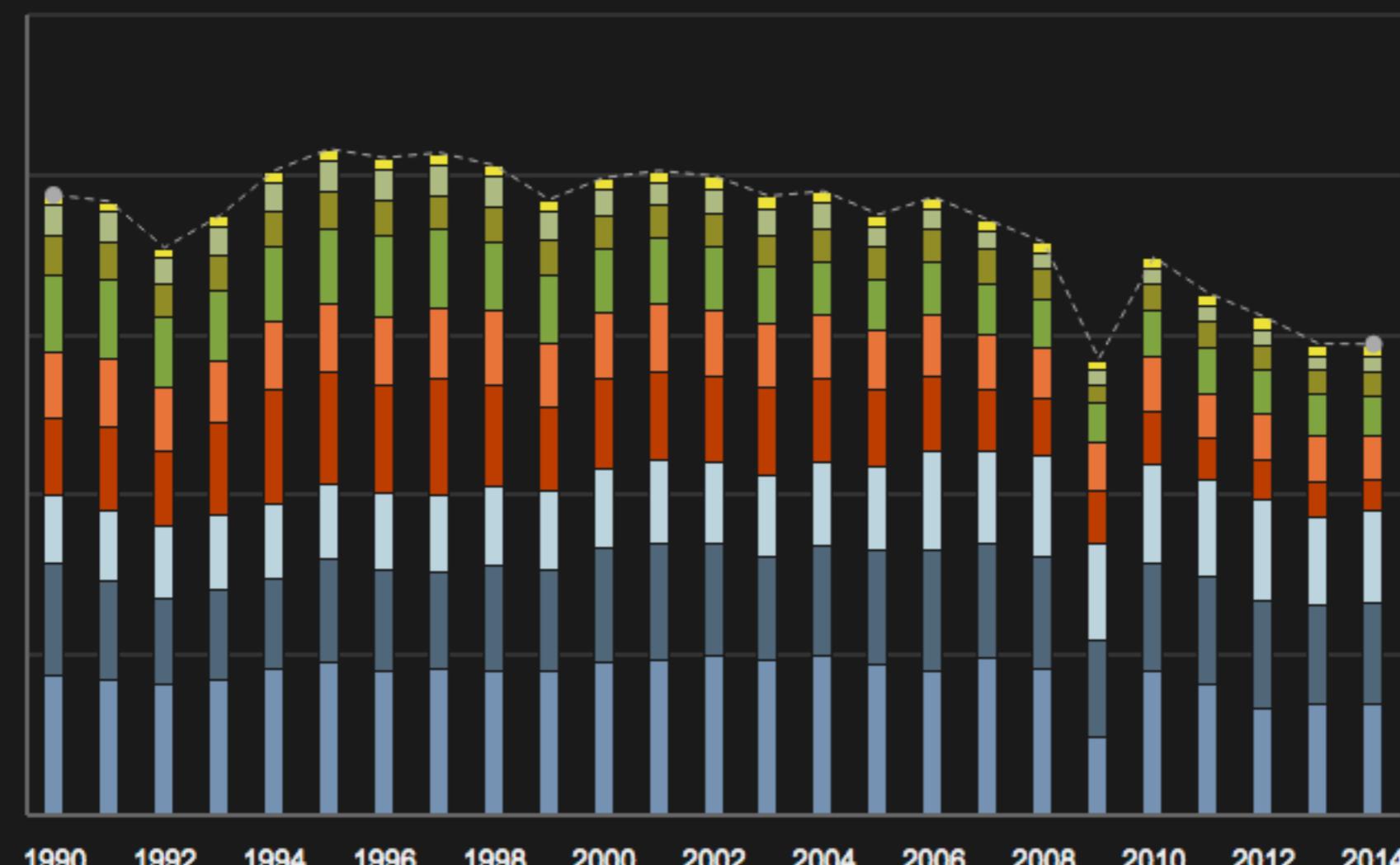
20000

15000

10000

5000

0



Utsläpp av växthusgaser från Industrin 1990–2014

Utsläpp av växthusgaser från Industrin

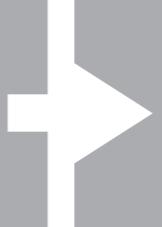
Tillhör Sveriges officiella statistik.

Dölj alla

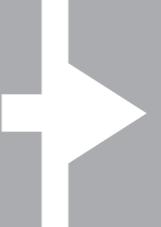
- Koksverk
- Livsmedelsindustri
- Metallindustri (exkl. järn- och stål)
- Övrig (gruvindustri, trävaruhandel m.m.)
- Kemiindustri
- Pappers- och massa-industri samt tryckerier
- Raffinaderier samt distribution av olja och gas
- Mineralindustri (exkl. metaller)
- Järn- och stålindustri
- Totalt



Why?



What?



How?



Do!

- From ideas and sketches to **production**.
- What programs and techniques do we need to deliver?



Overview Examples Documentation Source

DB Data-Driven Documents

CARTOON INDUSTRIES GALLERY HELP PRICING ENTERPRISE BLOG SIGN UP LOGIN

el-report FOLLOW Search Resource Governance Index and other pages EDIT PAGE

BASHBOARD EXPLORE NEW RESOURCE GOVERNANCE... SHARE

Home > Resource Governance Index > Resources and governance

Resources and governance

Collection Resource Governance Index

Only three of the world's most resource dependent countries rank above average on NRGI's rule of law index.

Datawrapper

An open source tool helping anyone to create simple, correct and embeddable charts in minutes.

Login / Sign Up Try it now!

Easy to use

This little tool reduces the time needed to create a correct chart and embed it into any

Full control

Datawrapper is 100% free software. You can use our free hosted service, or install it on

We encourage you to use a responsive layout.

Merge+ (revenuewatch_extractive_incomes_and_governance)

Connected to Excel

Workbook revenuewatch_extractive_incomes_a...

Sheets

Enter sheet name

- Full data
- HDI-Indicators
- Key data
- Merge
- WB_corruption

Merge Go to Worksheet

Country code (HDI-Indicators) Country (HDI) Country Resource Governance Index Pop

Merge	Merge	Merge	Merge
Nigeria	42		
Zambia	61		
Yemen	43		
Gabon	46		
Kazakhstan	57		
Bolivia	53		
Mongolia	51		

Contact us!

We're not around, but we'd love to chat another time. Click here and type your Name. Click here and type your Email. We're not around but we still want to hear from you! Leave us a note.

→ The simplest data import or export to help you to create visualizations.

→ Make sense of your location data and power your business.

→ Create the most beautiful visualizations on the web with Datawrapper.

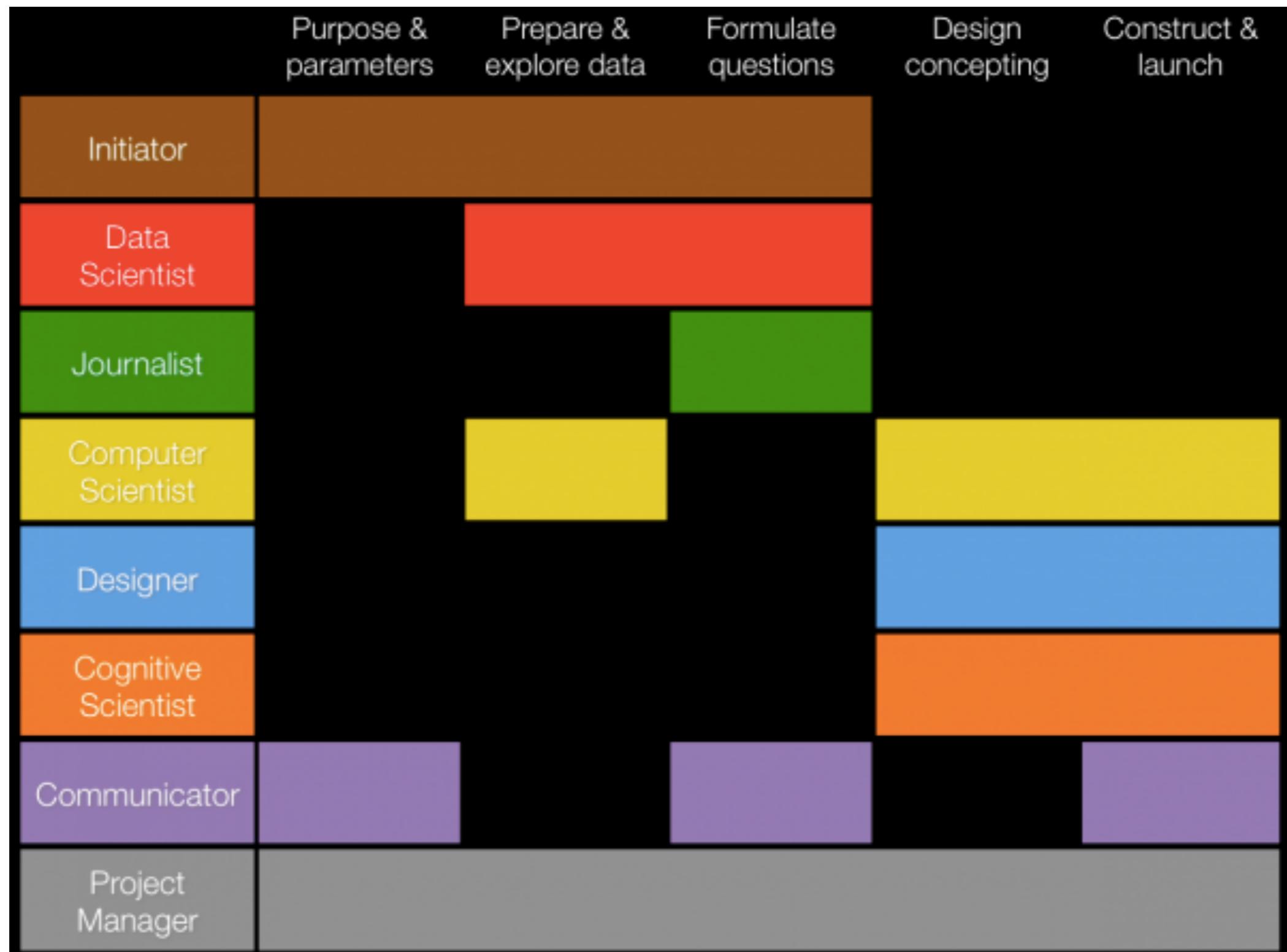
ODYSSEY

A simple way for journalists, designers, and creators to weave interactive stories

CREATE STORY DOCUMENTATION

Odyssey is a product of CARTOON, the most powerful mapping tool out there.

@jensthnas



@jensfinnas



The course.

@jensfinnas



14/12

Introduction and Excel exercises.

17/12

Workshop about finding and researching the story in data with **Paloma Pérez Lucero** from TT.

18/1

Visualizing lots of data on small screens with **Henrik Hedvall** from Tink.

20/1

Workshop about visualizing data with **Information is beautiful**.

22/1

Mid-term presentations

4/2

Final presentations



The assignment

Analyze a set of data.

Find a story.

Present it.



Clients

Interview the client.

Interview the data.

@jensfinnas



Final product

Should be digital and presented online.

A interactive application, an infographic, a video.

Or something else?



Prepare to answer

What do you want to say?

What is the purpose?

Who is the audience?

Why did you choose this form?



Evaluation

Is the story interesting? And does the data support it?

Do we understand the visualization?
Is it aesthetic and well designed?



Init		
 jensfinnas authored 31 minutes ago		latest commit <code>1da1f33408</code> 
.gitignore	Init	31 minutes ago
README.md	Init	31 minutes ago

[README.md](#)

DDS: Data visualization and storytelling

The assignment

In short: analyze a set of data, find a story, present it.

What data?

You will be provided with a couple of datasets from actual clients, but you are free to choose another set of data if you want. If you do so, verify that it meets the course requirements with Jens Finnäs and Thomas Björk before you proceed. You are expected **analyze the data yourselves** and draw your own conclusions from it.

@jensfinnas