The secret to good data visualization

…and why we have this course.

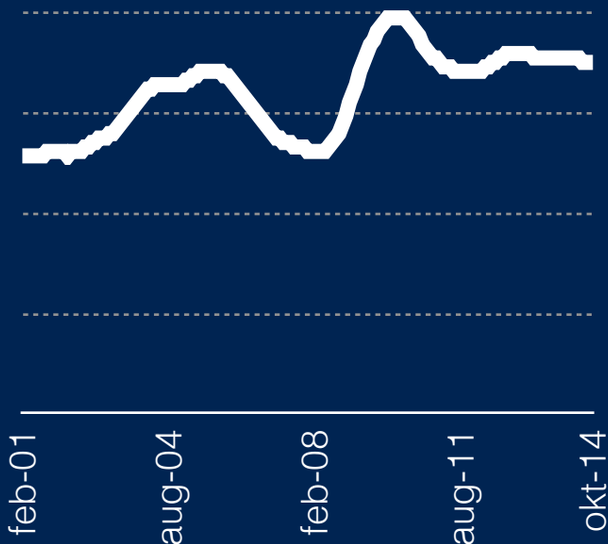
@jensfinnas

To **understand society** you have to **understand data**.

@jensfinnas

World ⇨ Data ⇨ Visualization

	A	B	C
1	**Arbetslöshet**		
2	Andel av arbetskraften, 15-74 år		
3	**Månad**	**Procent**	**Procent, säson**
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

So what is the secret

good data visualization?

Coding for Journalists 101 : A four-part series

Tweet 20

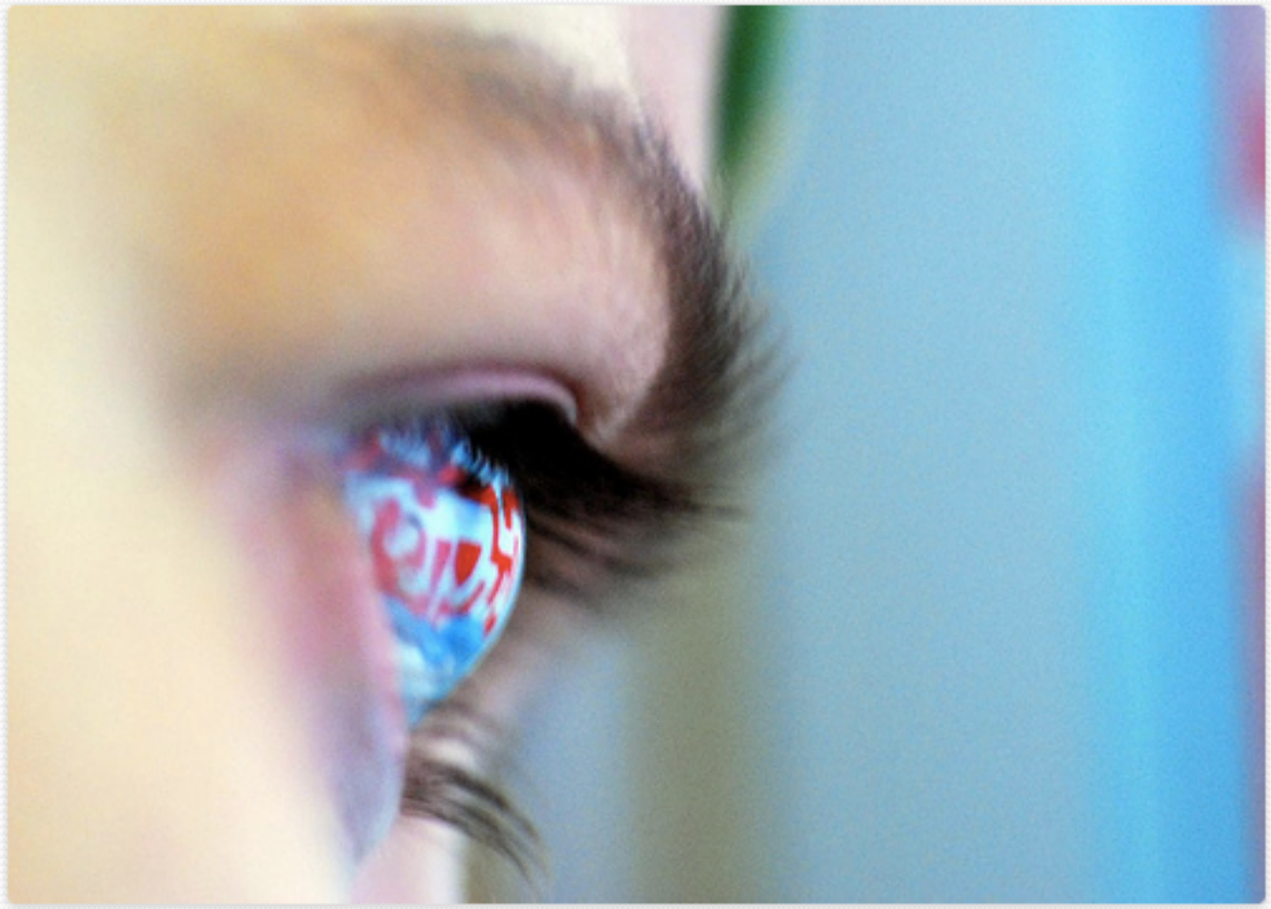


Photo by Nico Cavallotto on Flickr

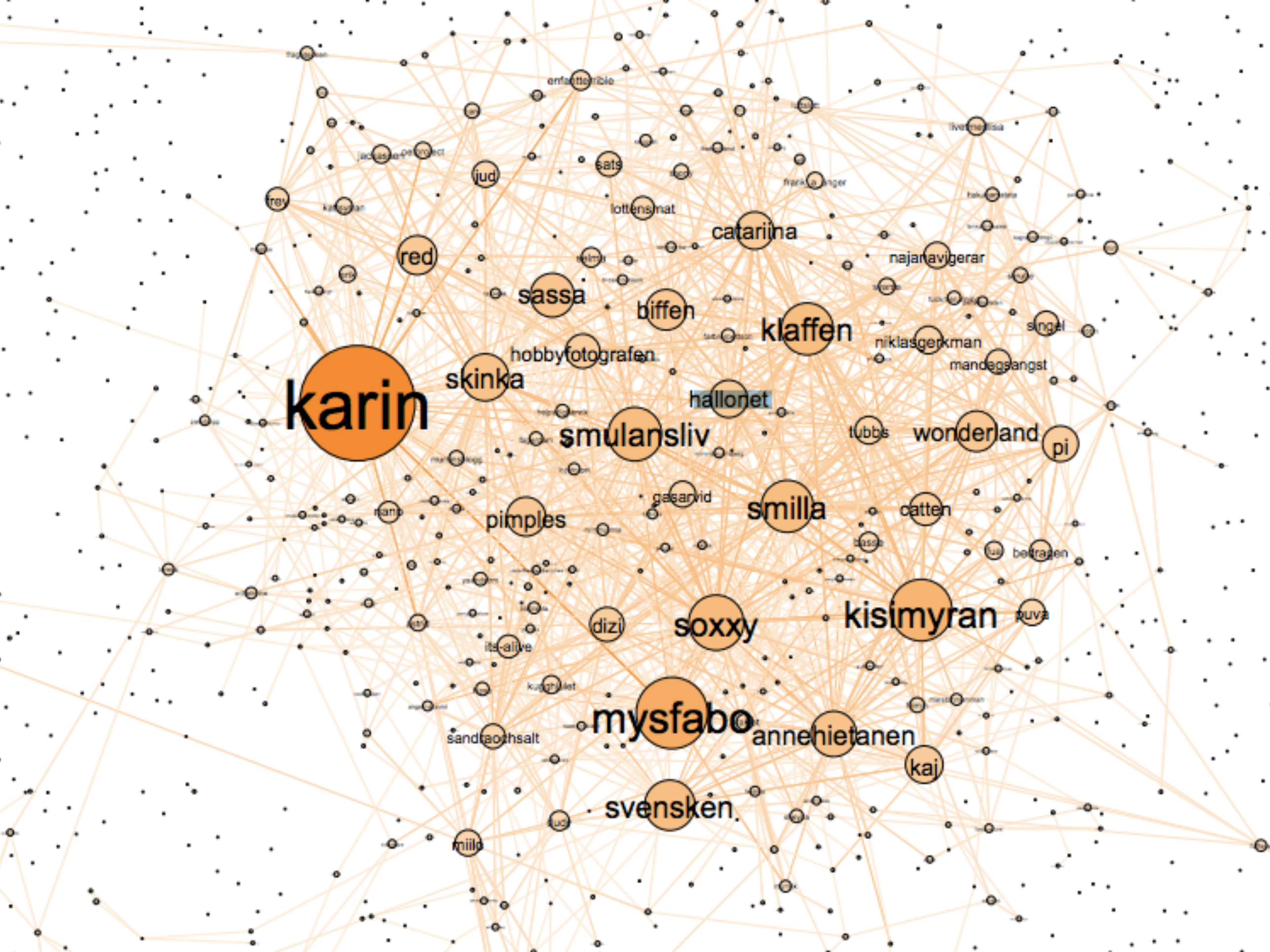
```ruby
require 'rubygems'
 require 'nokogiri'
 require 'open-uri'

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://pussan.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


karin
mysfabo
smulansliv
svensken
kisimyran
soxxy
smilla
pimples
catariina
sassa
red
biffen
klaffen
skinka
hobbyfotografen
hallonet
wonderland
pi
annehietanen
kaj
dizi
sats
jud
lottensmat
najanavigerar
niklasgerkman
mandagsangst
singel
tubbs
catten
gasarvid
its-alive
sandraothsalt
miild

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.

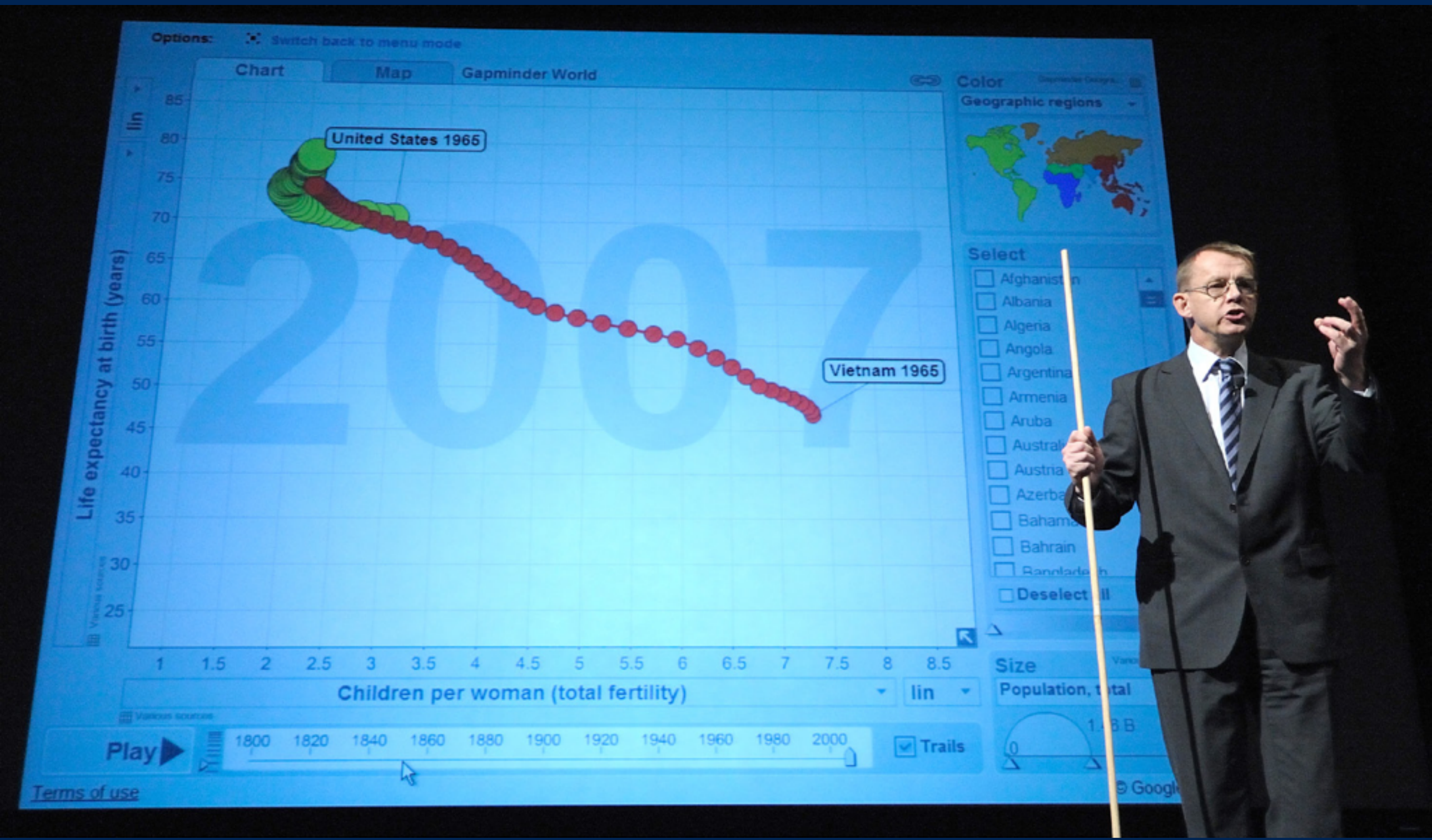
The author is **Jens Finnäs**, a freelance journalist from Finland, currently living in Stockholm, Sweden.

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**


Options:
Switch back to menu mode
Chart
Map
Gapminder World
United States 1965
Vietnam 1965
2007
Life expectancy at birth (years)
lin
85
80
75
70
65
60
55
50
45
40
35
30
25
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
Children per woman (total fertility)
lin
Various sources
Play
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000
Trails
Terms of use
Color
Geographic regions
Select
Afghanistan
Albania
Algeria
Angola
Argentina
Armenia
Aruba
Australia
Austria
Azerbaijan
Bahamas
Bahrain
Bangladesh
Deselect all
Size
Population, total
0
1.3 B
Google

@jensfinnas

What makes Hans Rosling's data visualizations great?

@jensfinnas

Medium

How is **digital different than analogue** as a data viz medium?

@jensfinnas

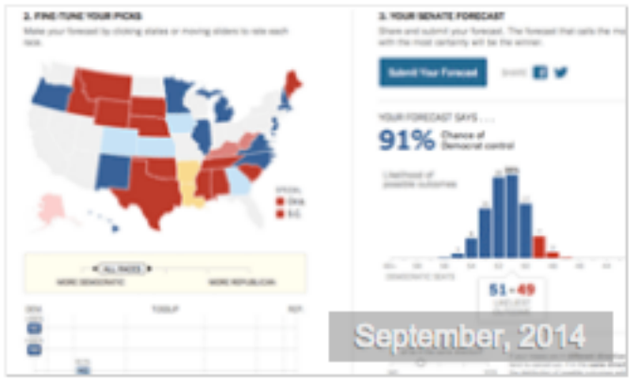
1. Animation.
2. Interaction.
3. Narration.
4. Boundlessness.

@jensfinnas

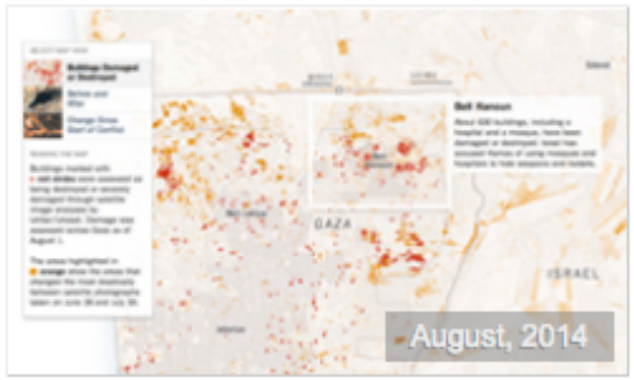
Heavier Babies Do Better in School

Flooding Risk From Climate Change, Country by Country

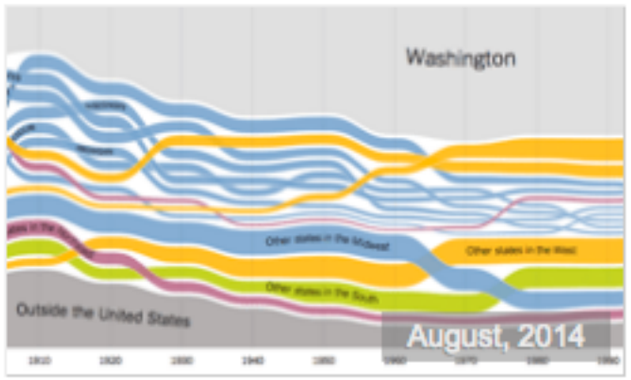
How Likely Is It That Birth Control Could Let You Down?



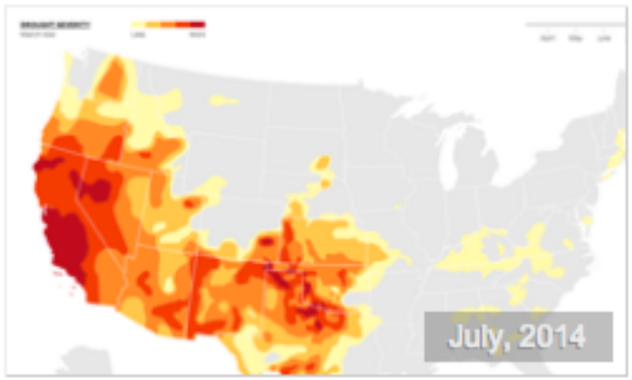
Elections 2014: Make Your Own Senate Forecast



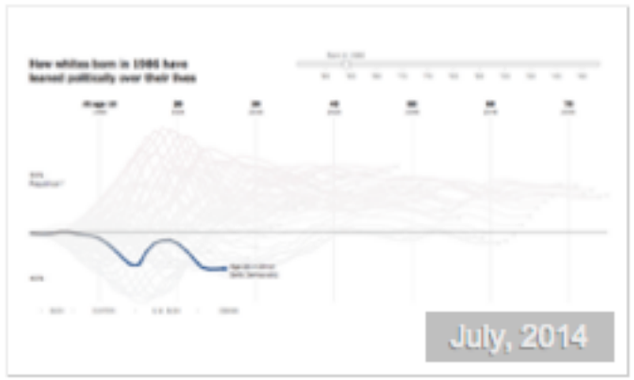
Assessing the Damage and Destruction in Gaza



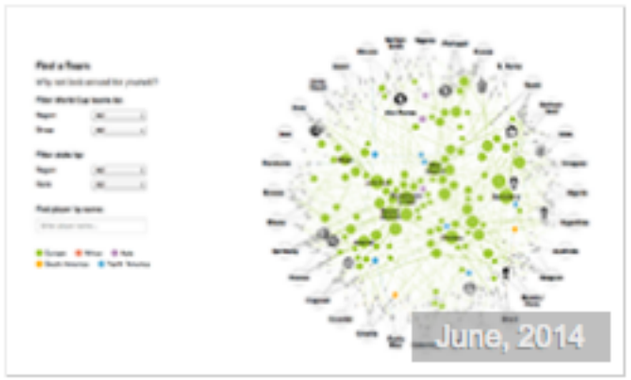
Where We Came From, State by State



Mapping the Spread of Drought Across the U.S.



How Birth Year Influences Political Views



The Clubs That Connect the World Cup

http://collection.marijerooze.nl/

@jensfinnas

Medium

Medium

Sender ⇨ **Message** ⇨ Receiver

Good data visualization

tells a story.

@jensfinnas

Good data visualization

has something to say.

@jensfinnas

How to do it?

Find the
story

@jensfinnas

Start with **a question**
or with **set of data**.

@jensfinnas

What do you **want to say**?

@jensfinnas

Is the purpose to…

…**explain** a phenomenon?

…make **an argument**?

…let the **user explore**?

@jensfinnas

Who is your **audience**?

Find the story

Choose the form

@jensfinnas

Do you want to highlight…
…a **difference**?
…a **change over time**?
…a **relationship**?
…a **geographical pattern**?

Find the story

Choose the form

Visualize

@jensfinnas


Overview
Examples
Documentation
Source
Data-Driven Documents

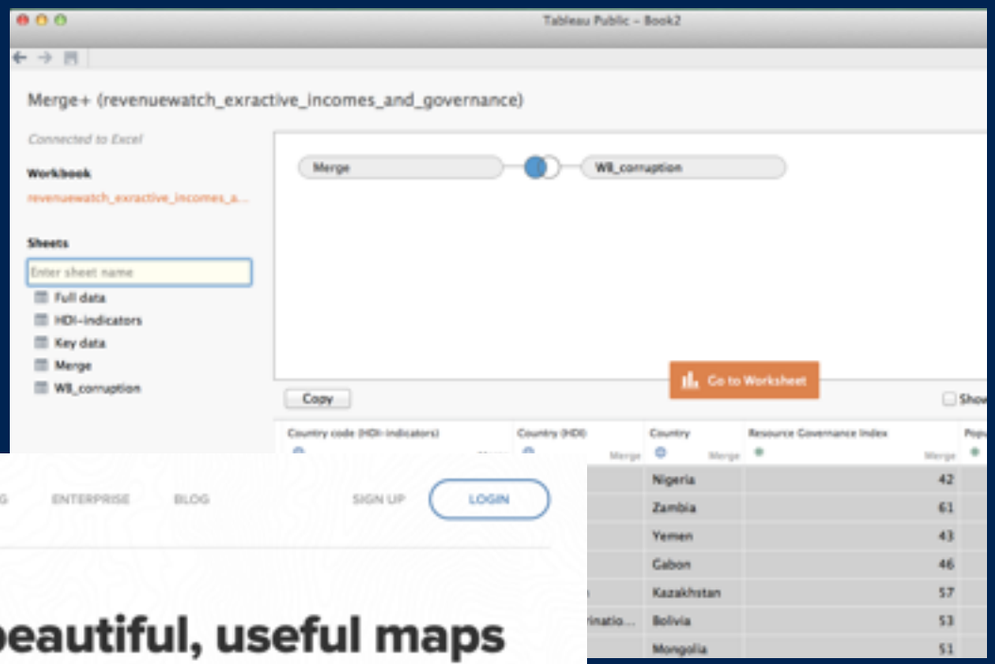
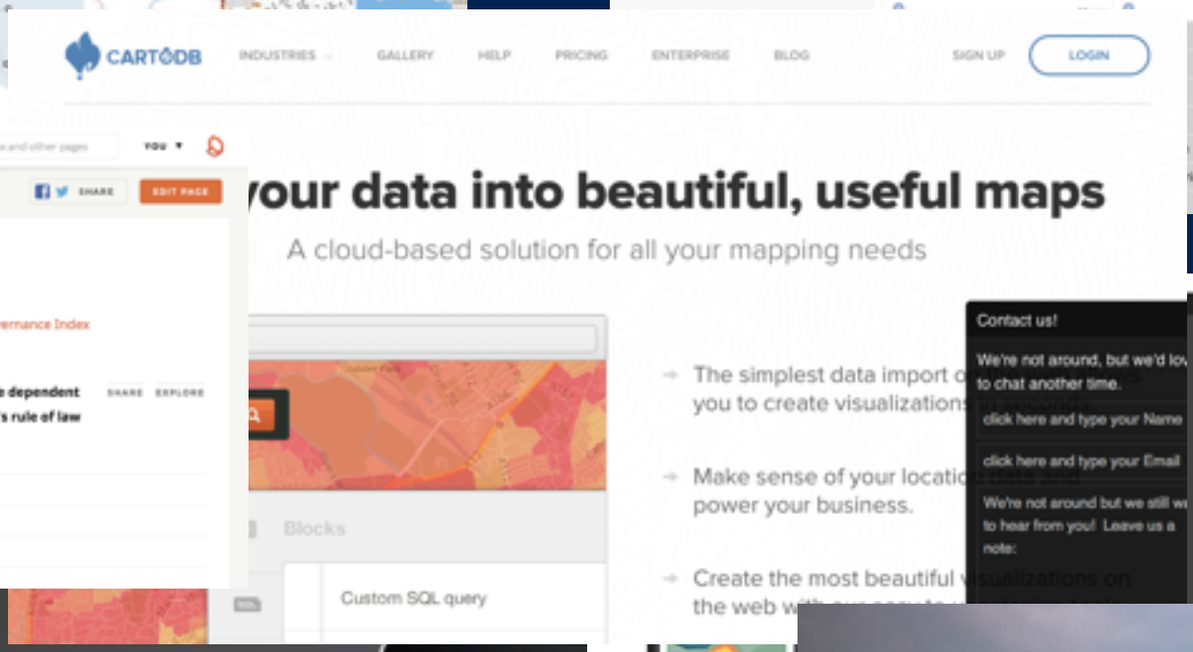
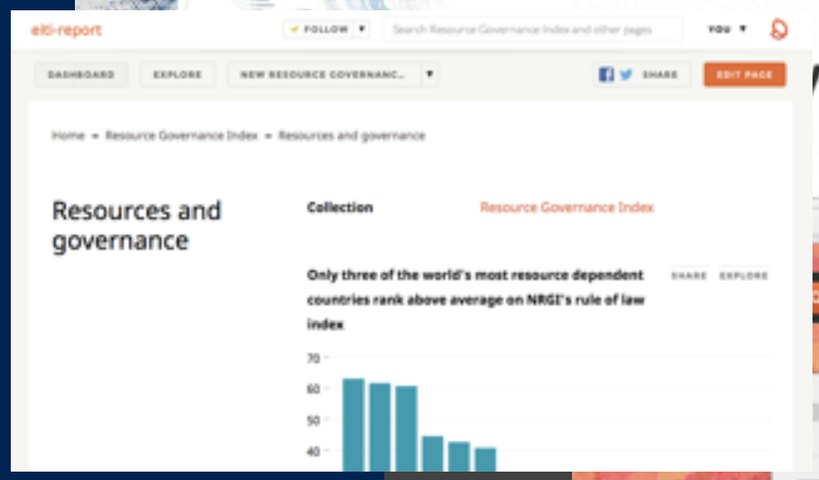
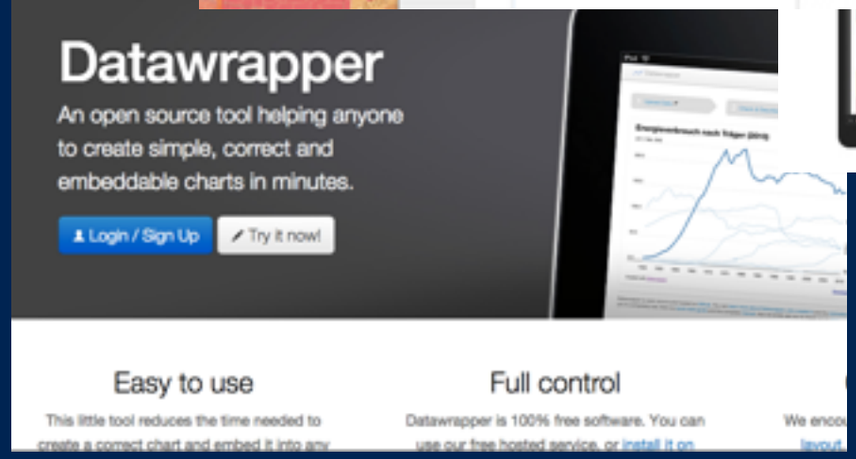


Tableau Public – Book2
Merge+ (revenuewatch_exractive_incomes_and_governance)
Connected to Excel
Workbook
Sheets
Full data
HDI-indicators
Key data
Merge
WB_corruption
Copy
Go to Worksheet
Country code (HDI-indicators)
Country (HDI)
Country
Resource Governance Index
Nigeria 42
Zambia 61
Yemen 43
Gabon 46
Kazakhstan 57
Bolivia 53
Mongolia 51


CARTODB
INDUSTRIES
GALLERY
HELP
PRICING
ENTERPRISE
BLOG
SIGN UP
LOGIN
your data into beautiful, useful maps
A cloud-based solution for all your mapping needs
Blocks
Custom SQL query
Contact us!

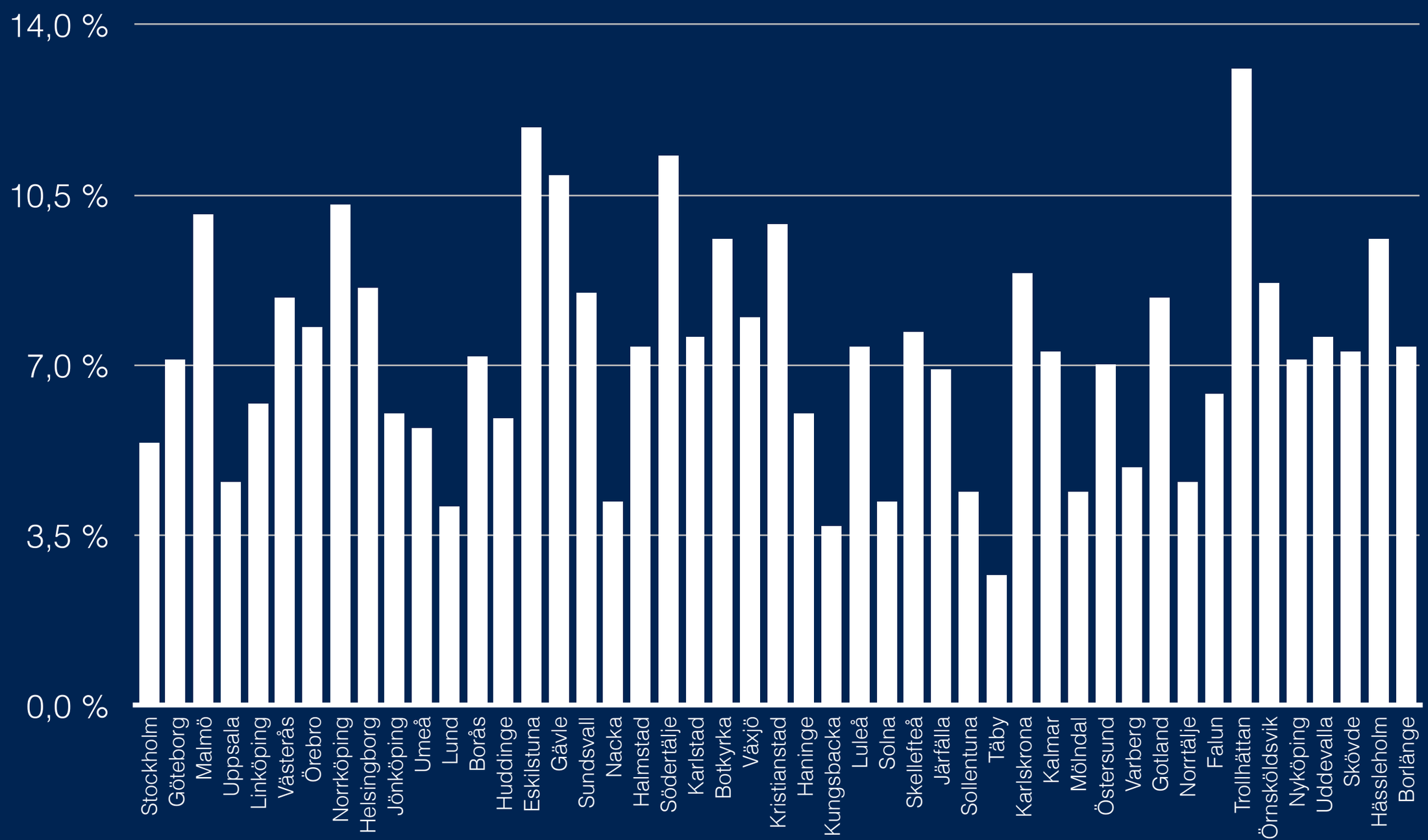

eiti-report
DASHBOARD
EXPLORE
NEW RESOURCE GOVERNANC...
SHARE
EDIT PAGE
Home » Resource Governance Index » Resources and governance
Resources and governance
Collection
Resource Governance Index
Only three of the world's most resource dependent countries rank above average on NRGI's rule of law index


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
Full control
Datawrapper is 100% free software. You can


ODYSSEY
A simple way for journalists, designers, and creators to weave interactive stories
CREATE STORY
DOCUMENTATION
Odyssey is a product of CARTODB, the most powerful mapping tool out there.

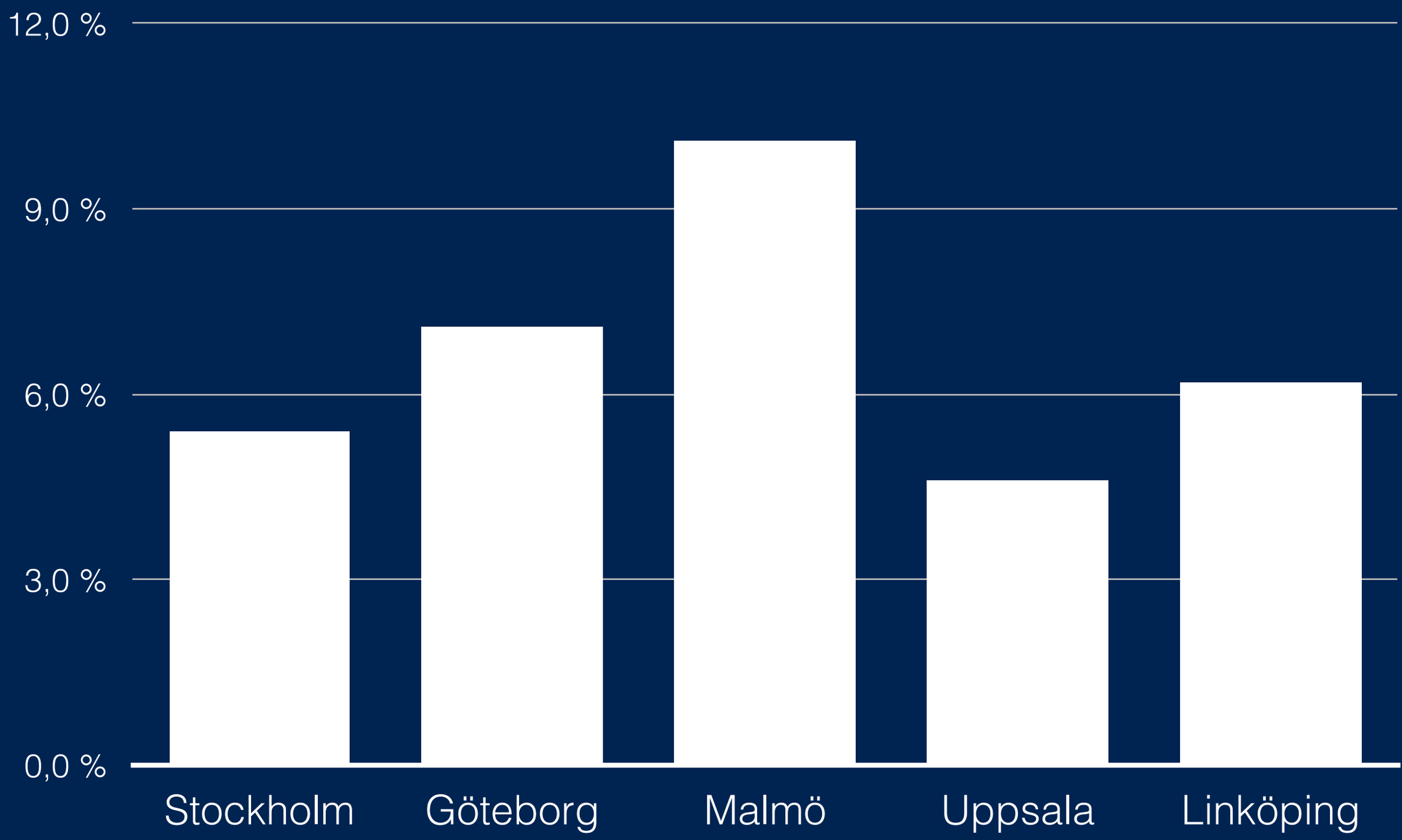
@jensfinnas

Unemployment 2013



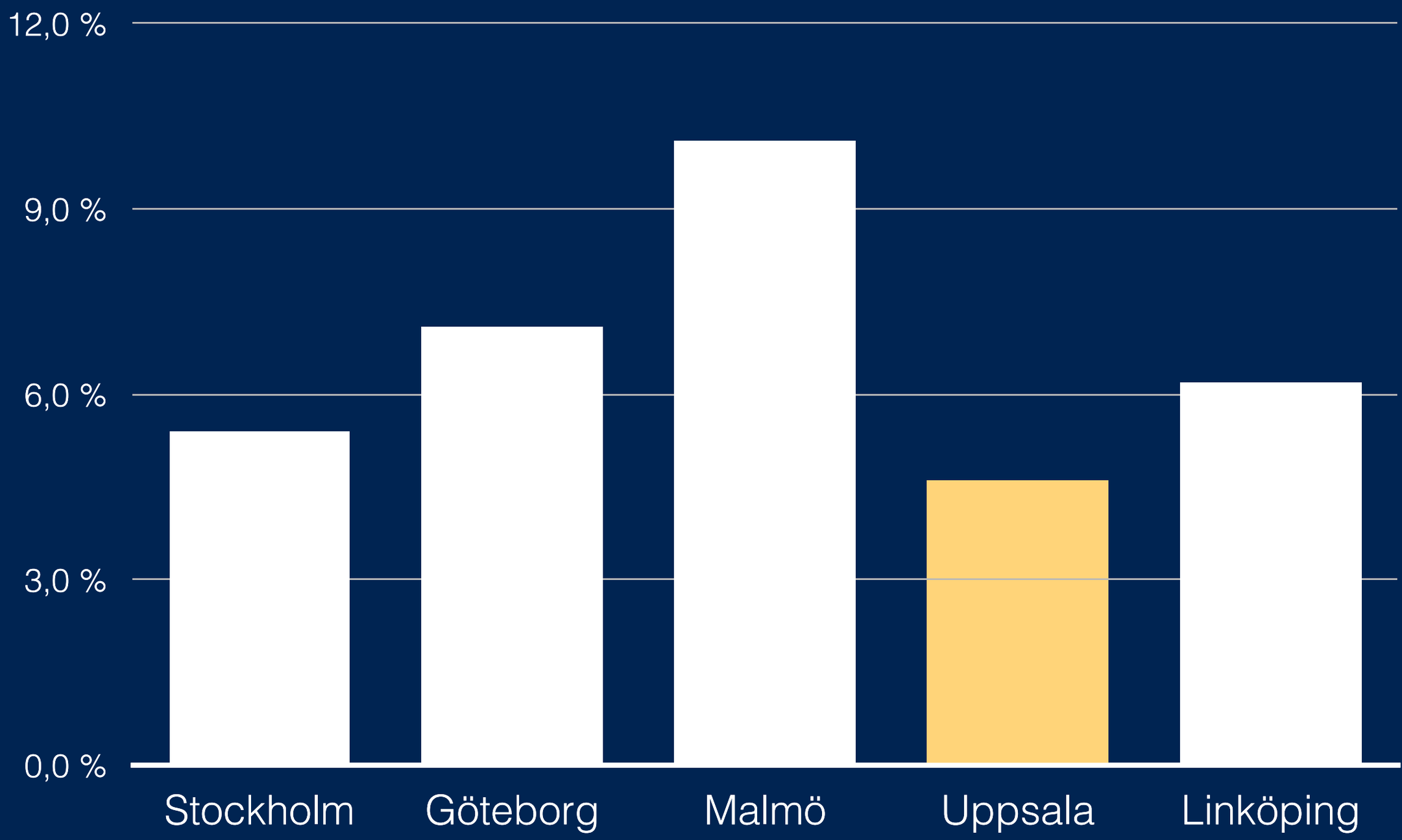
@jensfinnas

Unemployment 2013



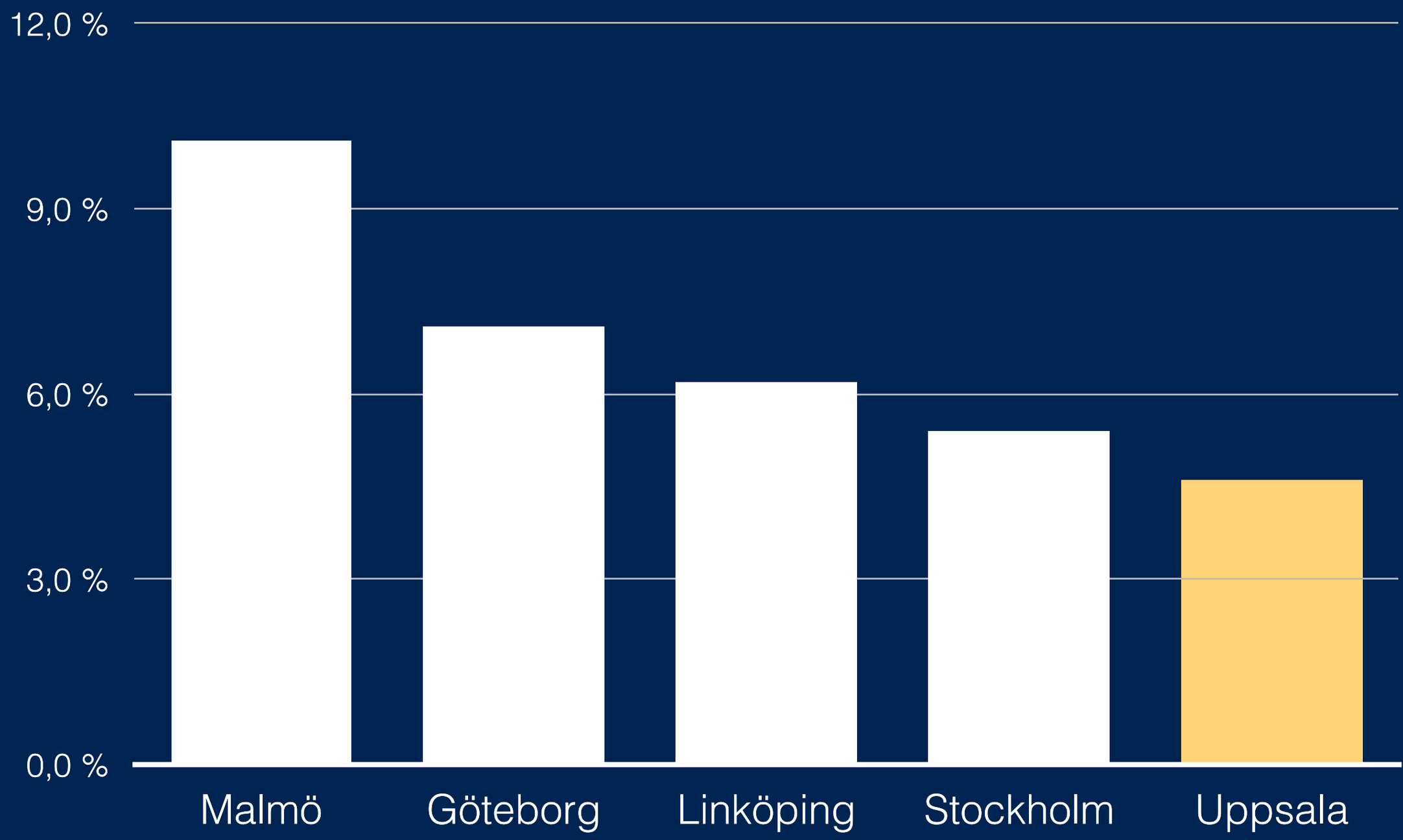
@jensfinnas

Arbetslöshet 2013



@jensfinnas

Arbetslöshet 2013

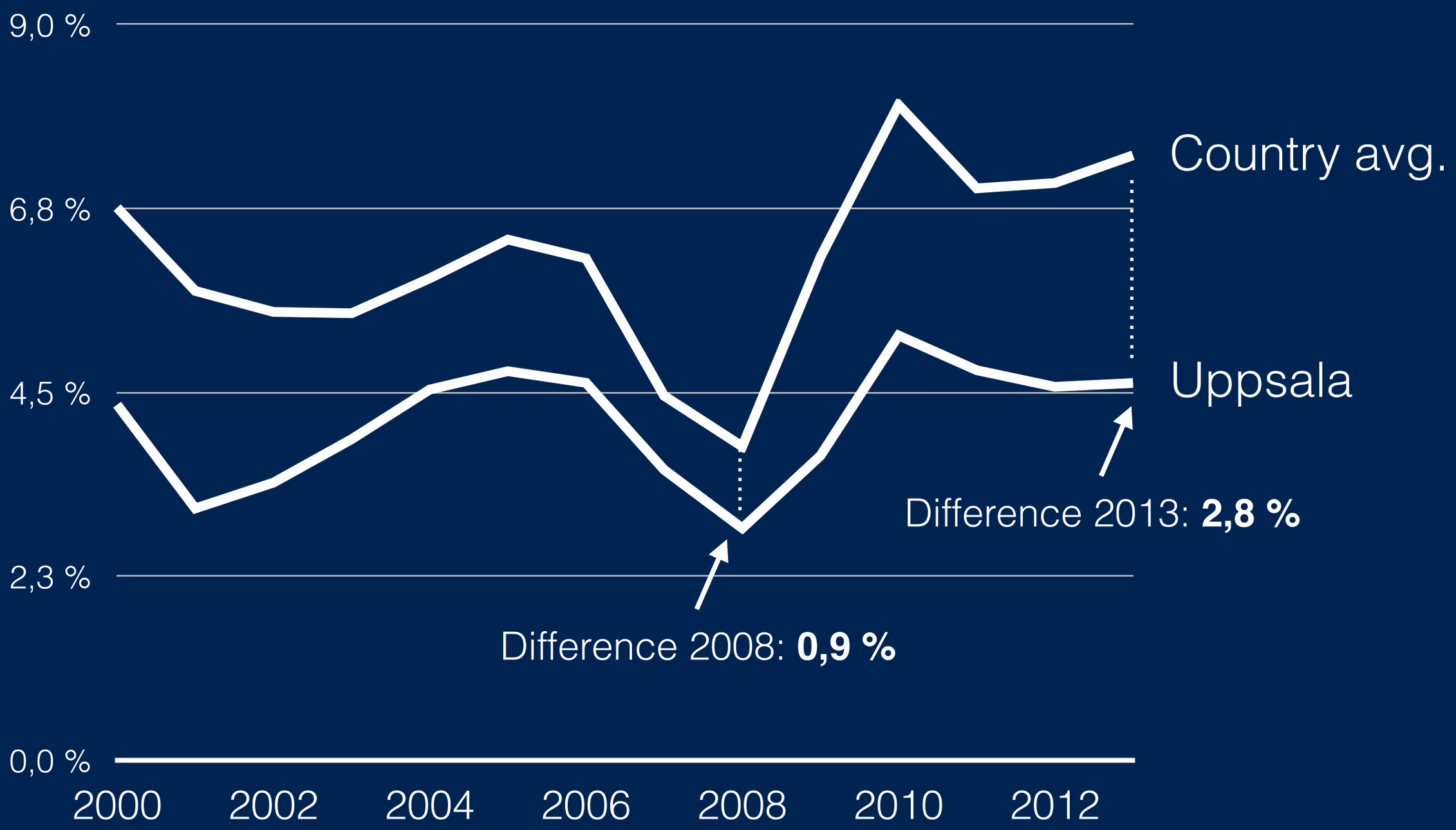


@jensfinnas

Unemployment 2000-2013

@jensfinnas

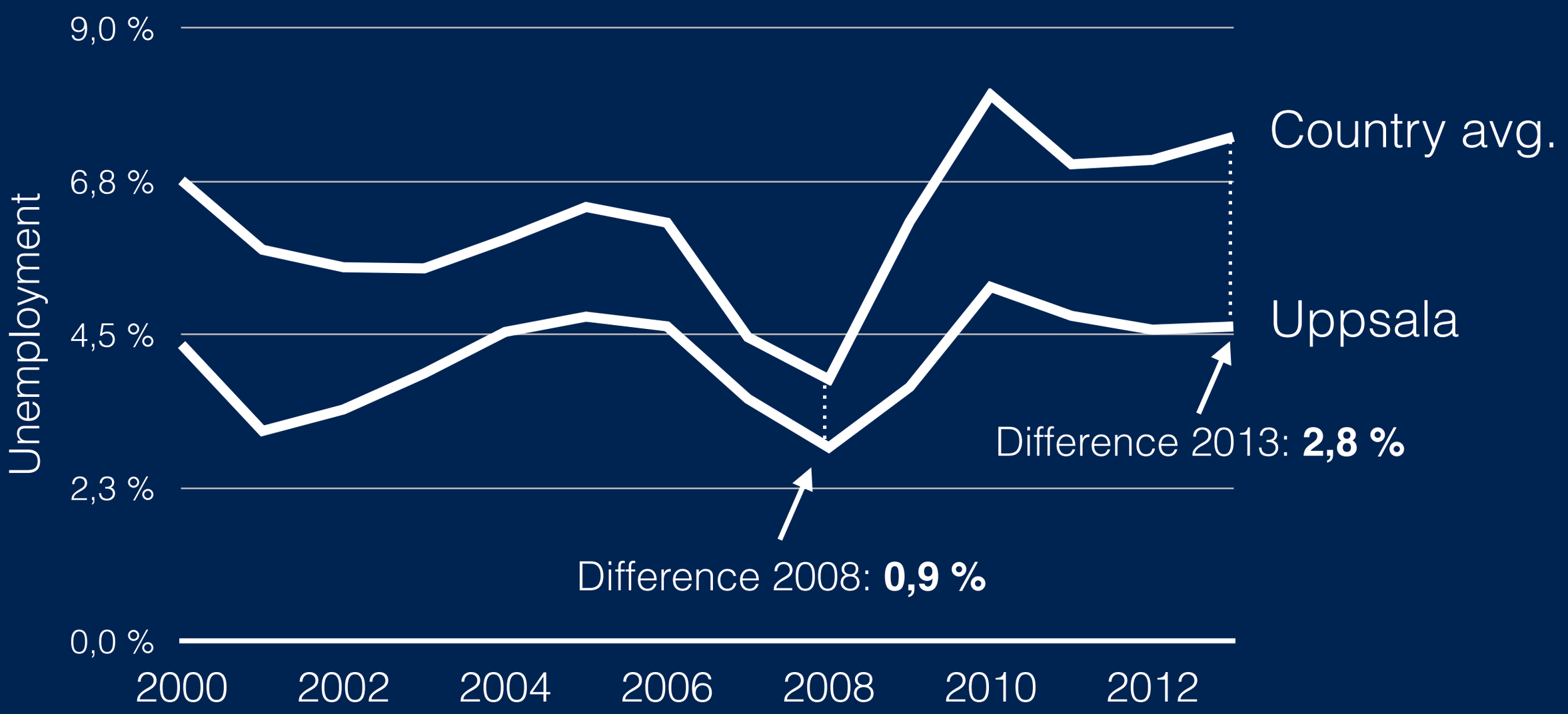
Unemployment 2000-2013



@jensfinnas

Uppsala recovered the crisis better

Since 2008 unemployment has risen, but not as much as in the rest of Sweden.



@jensfinnas

Unemployment 2000-2013

@jensfinnas

Title
Annotation
Filtering
Coloring
Sorting

@jensfinnas

Find the story

Choose the form

Visualize

@jensfinnas

The course.

@jensfinnas

8.12 Introduction

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

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

12.12 Excel workshop and data viz tools with **Jens Finnäs**.

15.12 Visualizing data online with **Daniel Lapidus**.

14.1 Mid-term presentations

15.1 Canvas magic and HTML5 with **Cas Lemmens**.

29.1 Final presentaions

@jensfinnas

The assignment

Analyze a set of data.

Find a story.

Present it.

@jensfinnas

Data

Use a client dataset or find your own.

Final product

A interactive application, an infographic, a video.
Or something else?

@jensfinnas

Presentations

Mid-term: January 15th

Online: January 27th

Final: January 29th

@jensfinnas

Prepare to answer
What do you want to say?
What is the purpose?
Who is the audience?
Why did you choose this form?

@jensfinnas

Evaluation

Is the story interesting?

Do we understand it?

Is it aesthetic and well designed?

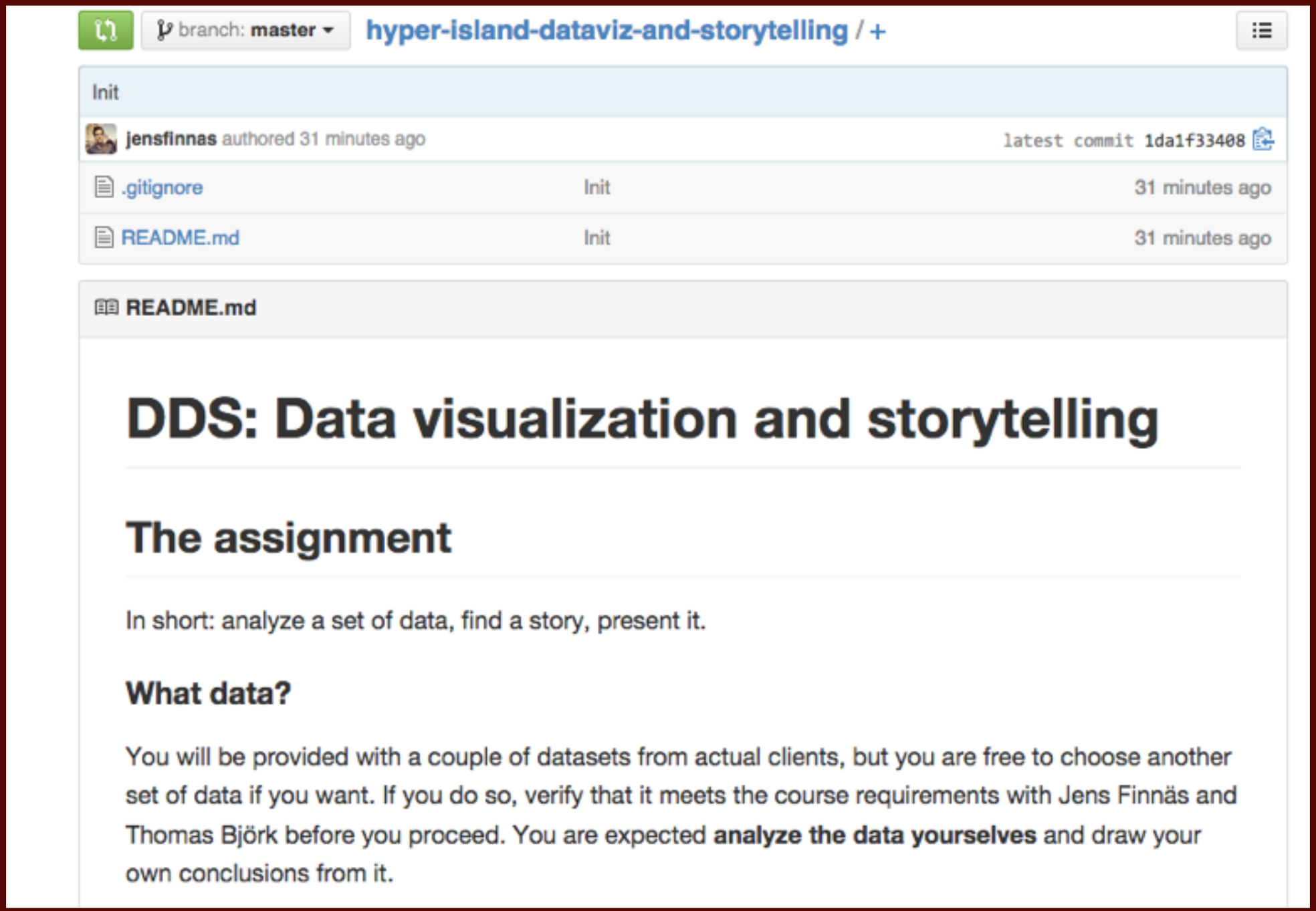
Find the story

Choose the form

Visualize

@jensfinnas

bit.ly/dds-2014



@jensfinnas