

DATA SCIENCE SHOW & TELL

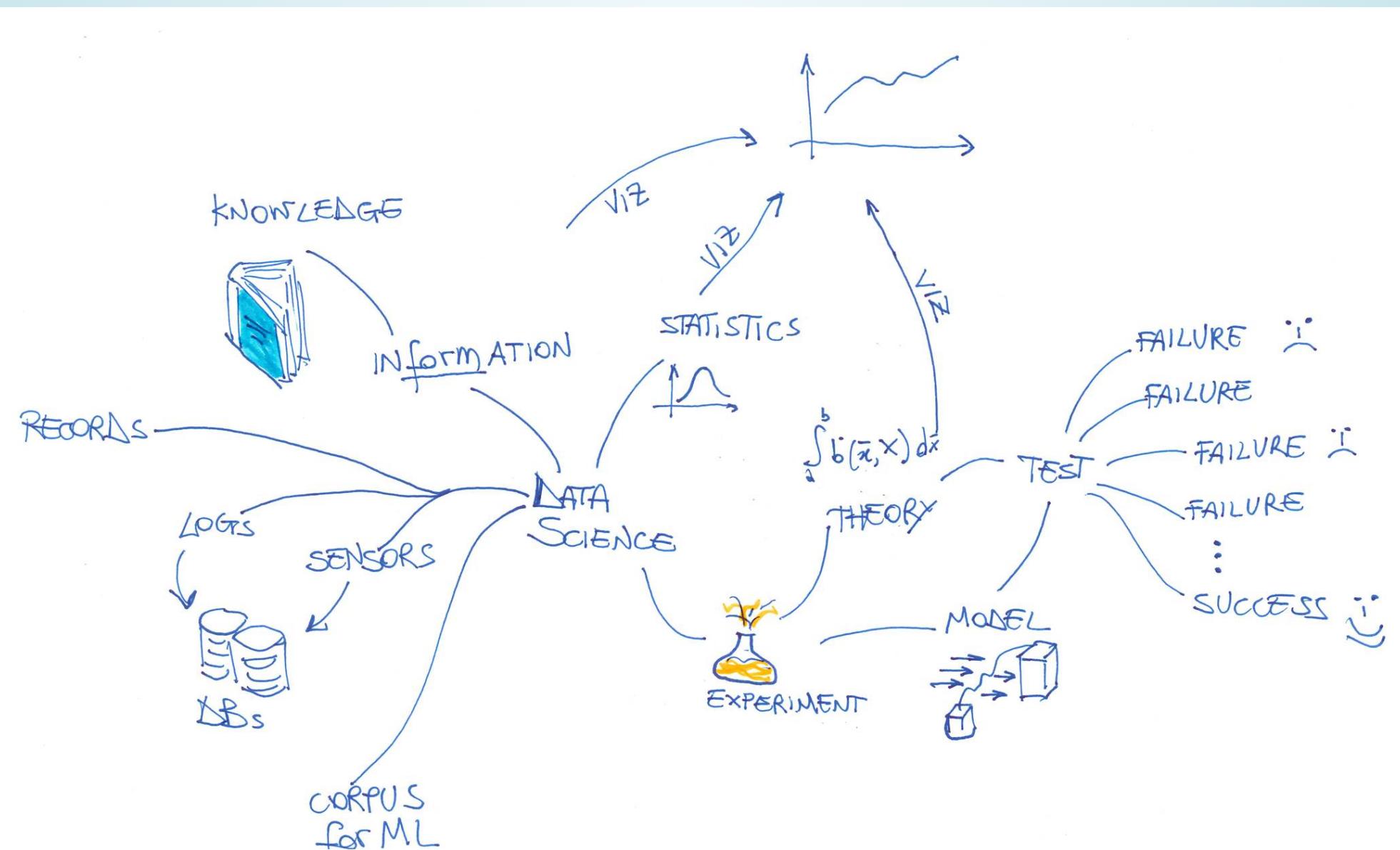
ENRICO SPINIELLI

JUNE 9, 2016

Live slides available at <https://espinelli.github.com/showandtell>

PDF and source of slides available at
<https://github.com/espinelli/showandtell>





GIVEN:

- life is *short*
- I am lazy
- You should not lie
- Humans are intelligent (w/ caveats ;-)
- ... and not all of them are working at Eurocontrol

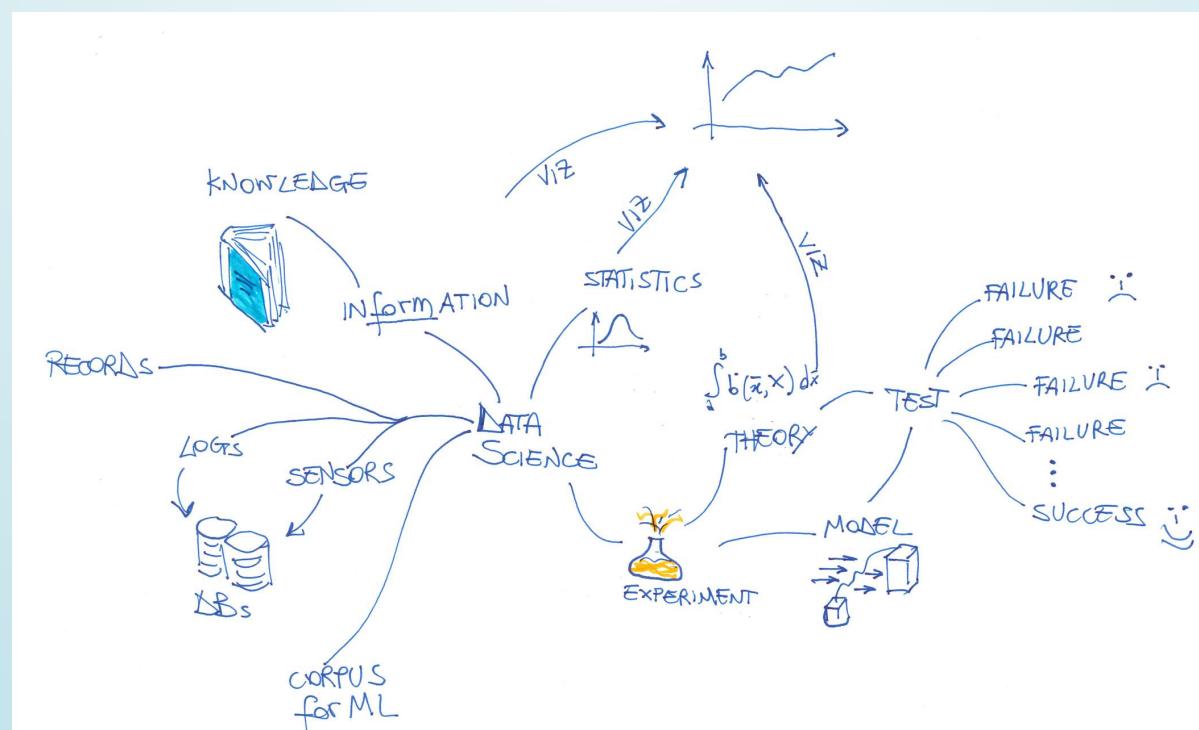
...IT FOLLOWS

- I'll (procrastinate on boring stuff and only) work on useful/fun projects
- Automation saves me from repeating boring and/or forgotten tasks
- I'll be open to let others critisize/scrutinize/learn
- ...and I'll learn back from them
- I'll strive to produce truthful explanations/visualizations

LET'S DO IT!

THE AXIOMS (IMHO)

- Value of data --> visualization
- Visualization --> WWW
- Make *data* available



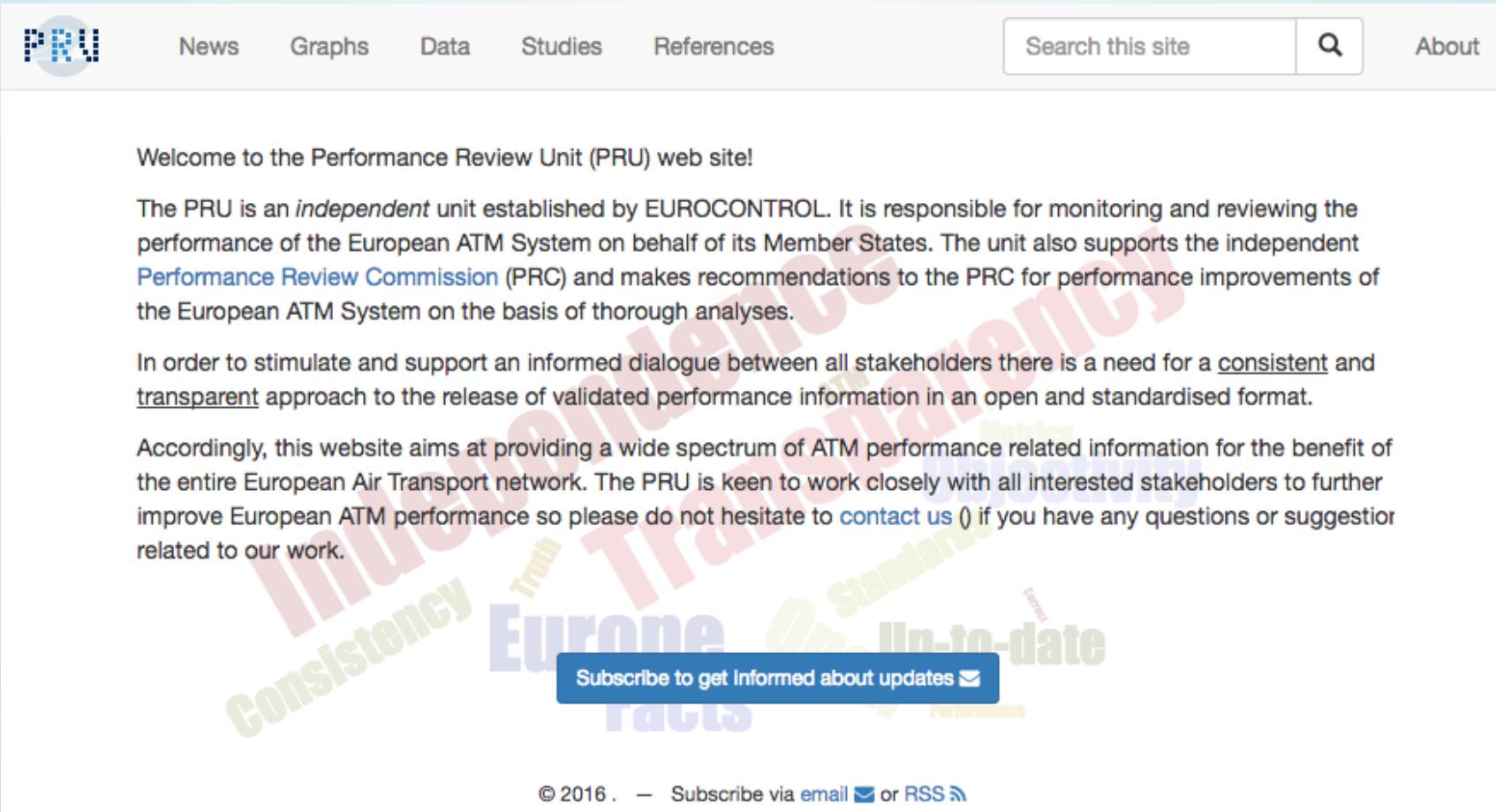
THE PLAN

(JAN 2015)

- Generate a (static) website for the PRU
- Version control it all
- Automate!

NOW

ONE YEAR AND A HALF LATER



The screenshot shows the PRU website homepage. At the top, there is a navigation bar with links for News, Graphs, Data, Studies, References, a search bar, and an About link. The main content area starts with a welcome message: "Welcome to the Performance Review Unit (PRU) web site! The PRU is an *independent* unit established by EUROCONTROL. It is responsible for monitoring and reviewing the performance of the European ATM System on behalf of its Member States. The unit also supports the independent Performance Review Commission (PRC) and makes recommendations to the PRC for performance improvements of the European ATM System on the basis of thorough analyses." Below this, there is a statement about the website's purpose: "In order to stimulate and support an informed dialogue between all stakeholders there is a need for a consistent and transparent approach to the release of validated performance information in an open and standardised format." Further down, there is a statement about the website's mission: "Accordingly, this website aims at providing a wide spectrum of ATM performance related information for the benefit of the entire European Air Transport network. The PRU is keen to work closely with all interested stakeholders to further improve European ATM performance so please do not hesitate to [contact us](#) if you have any questions or suggestions related to our work." At the bottom of the page, there is a call-to-action button that says "Subscribe to get Informed about updates" with an envelope icon.

Welcome to the Performance Review Unit (PRU) web site!

The PRU is an *independent* unit established by EUROCONTROL. It is responsible for monitoring and reviewing the performance of the European ATM System on behalf of its Member States. The unit also supports the independent Performance Review Commission (PRC) and makes recommendations to the PRC for performance improvements of the European ATM System on the basis of thorough analyses.

In order to stimulate and support an informed dialogue between all stakeholders there is a need for a consistent and transparent approach to the release of validated performance information in an open and standardised format.

Accordingly, this website aims at providing a wide spectrum of ATM performance related information for the benefit of the entire European Air Transport network. The PRU is keen to work closely with all interested stakeholders to further improve European ATM performance so please do not hesitate to [contact us](#) if you have any questions or suggestions related to our work.

Subscribe to get Informed about updates

© 2016. — Subscribe via [email](#) or [RSS](#)

SECTIONS

[News](#)[Graphs](#)[Data](#)[Studies](#)[References](#)[About](#)

DEMO TIME

- Make sure to navigate from Data to Metadata (last column in the table)
- Check the Graphs out
- Enjoy the Studies

the official PRU site, <http://ansperformance.eu>

FEATURES



euctrl-pru / website

Code Issues 14 Pull requests 1 Wiki Pulse Graphs Settings

Unwatch 4 Star 0 Fork 0

the source of the web site <https://euctrl-pru.github.io/> — Edit

319 commits 2 branches 0 releases 3 contributors

Branch: master New pull request New file Find file HTTPS https://github.com/euctrl Download ZIP

Commit	Message	Date
	espinelli filtered unnecessary FAB CE's; formatted Avg delay column.	Latest commit 8cbe312 6 days ago
	_assets filtered unnecessary FAB CE's; formatted Avg delay column.	6 days ago
	_data added generate footer	2 months ago
	_includes take 'Generated ...' away	2 months ago
	_layouts fix creation of year heading	13 days ago
	_plugins simpler rendering of definitions and anspes	3 months ago
	_posts fixt typo in RN and misnamed ids for tabs in graphs page	13 days ago
	_sass integrated dashboard as submodule; use bootstrap and adapted styles.	11 months ago
	_studies updated CSV files; fixed typo in intra-eu flows study	2 months ago
	_styles pandoc generated site.	a year ago
	about folded contact us in about	3 months ago
	css introduced assets management	7 months ago



<USER>.GITHUB.IO

This repository Search Pull requests Issues Gist

euctrl-pru / euctrl-pru.github.io Unwatch 1 Star 1 Fork 0

Code Issues 0 Pull requests 0 Wiki Pulse Graphs Settings

PRU public website <https://euctrl-pru.github.io> — Edit

40 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

espinelli Updating to euctrl-pru/website.github.io@35f4faa ... Latest commit 7e0bd6b 11 days ago

about	Updating to euctrl-pru/website.github.io@dc9b49e	14 days ago
assets	Updating to euctrl-pru/website.github.io@35f4faa	11 days ago
css	rebuilding site Thu Dec 3 13:30:02 CET 2015	6 months ago
data	Updating to euctrl-pru/website.github.io@35f4faa	11 days ago
graphics	Updating to euctrl-pru/website.github.io@914cd87	12 days ago
graphs	Updating to euctrl-pru/website.github.io@914cd87	12 days ago
news	Updating to euctrl-pru/website.github.io@dc9b49e	14 days ago
posts	Updating to euctrl-pru/website.github.io@dc9b49e	14 days ago
references	Updating to euctrl-pru/website.github.io@dc9b49e	14 days ago

TECH DOCS

Principles

We use a [Feature Branch Workflow](#): it is *simple* and it is *working* in real life.

For more complicated productions, release-like, we could adopt [GitFlow](#) (see also "A successfull Git brancing model" ... or [Intro to GitHub Flow](#) and [this blog](#))

Organization of the repositories

We have two repositories:

1. `website` : where we define the structure and content of the site. To do so we use the ([Jekyll](#)) framework.
2. `eucctrl-pru.github.io` : where the statically generated (from the Jekyll repo above) site is version controlled.

GitHub provides a service whereby a specially named repository is published on the Internet (see [GitHub Pages](#))

Workflows

We have 3 types of activities:

- development of next website release
- fix of currently published release
- publication of a (fix/update of a) release

Development happens in a branch named `<4-digit-year><2-digit-month>-release` , i.e. `201606-`



The logo of the European Organization for the Safety of Air Navigation (EUROCONTROL) features a stylized blue 'E' shape composed of three curved segments, with a small pencil icon at the top right. Below the graphic, the word "EUROCONTROL" is written in a bold, blue, sans-serif font.

Setup

- Rationale
- Tools Installation and Setup
- Intro to everything

Workflows

- Workflows
- Contributor Workflow
 - Workflow Initialization
 - Local Testing
- Maintainer Workflow
 - Workflow Initialization
 - Local Testing
- Integrator Workflow
- Publisher Workflow

Clone this wiki locally

<https://github.com/eucctrl-p> 

RELEASE

euctrl-pru / website

Unwatch 4 Star 0 Fork 0

Code Issues 15 Pull requests 1 Wiki Pulse Graphs Settings

201606 release #97

Open espinelli wants to merge 5 commits into master from 201606-release

Conversation 0 Commits 5 Files changed 17 +414 -643

espinelli commented 14 days ago • edited Performance Review Unit - Eurocontrol member +

H = high[er] priority
M = medium priority

- Data download updates H
- Definition and map of AUA, ANSP used for traffic and delay M
- Interactive graphs M (@espinelli)
 - radio button for year in En-route ATFM delays by FAB (FIR) choropleth.
 - ...

Backlog

- ASMA and Taxi-out dataset granularity down to daily M
- Issue ACE metadata page, see also issues #46 , #47

Labels None yet

Milestone No milestone

Assignees No one—assign yourself

Notifications
You're receiving notifications because you were mentioned.

1 participant

BUGS

euctrl-pru / website

Unwatch 4 Star 0 Fork 0

Code Issues 15 Pull requests 1 Wiki Pulse Graphs Settings

Filters is:issue is:open Labels Milestones New issue

<input type="checkbox"/> ① 15 Open	<input checked="" type="checkbox"/> 36 Closed	Author ▾	Labels ▾	Milestones ▾	Assignee ▾	Sort ▾
<input type="checkbox"/> ① review bib entry for Sam's paper to ICNS #98 opened 12 days ago by espinelli						
<input type="checkbox"/> ① javascript libraries: CDN and local hosting enhancement #92 opened on Apr 7 by espinelli						
<input type="checkbox"/> ① accented apt names incorrectly displayed #90 opened on Mar 18 by espinelli						
<input type="checkbox"/> ① Travis-CI deploy: make it better enhancement #89 opened on Mar 17 by espinelli						
<input type="checkbox"/> ① FAB names inconsistencies #86 opened on Feb 8 by espinelli						
<input type="checkbox"/> ① Info popup background does not work in IE #80 opened on Nov 19, 2015 by espinelli						
<input type="checkbox"/> ① introduces code (style) checks						

EDITING

- easy, i.e. textual (ASCII, no HTML): separate *content* from *style*
- nice Math (via [MathJax](#)):

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n$$

- bibliography: cite and style
- templates for different kind of pages (Definitions, list of ANSP's, RN's)

MARKDOWN



No need to edit in HTML: we (mainly) use [Markdown](#) (from [Pandoc](#))

```
## Methodology

[Horizontal en-route flight efficiency methodology](/r/m/hfe_pi.html)
is fully consistent with the Single European Sky (SES)
Performance Scheme [see {`% cite pru-hfe-pi --file aviation %`}].

## Column naming and types

### HFE data

{:.metatable}
| Column name | Src | Label | Column description | Example |
|---|---|---|---|---|
| YEAR | NM | YEAR | Reference year | 2014 |
| MONTH_NUM | NM | MONTH_NUM | Month (numeric) | 9 |
| MONTH_MON | NM | MONTH_MON | Month (3-letter code) | JAN |
```



BIBLIO

ATC Pre-Departure Delay (IATA code 89) Dataset

Data description

The ATC pre-departure delay is a proxy for ATC induced delays at the departure stand to avoid queuing at the departure runway.

The data is based on [IATA delay code 89](#) [see [\(IATA, 2012\)](#)] which, besides delays caused by local ATC constraints, also includes delays for other reasons which may introduce a certain level of bias. Work is in progress to allow for a better identification of the different causes.

Column naming and types

Column name	Data Source	Label	Description
YEAR	Network Manager	YEAR	Reference year
MONTH_NUM	Network Manager	MONTH_NUM	Month (numeric)
MONTH_MON	Network Manager	MONTH_MON	Month (3-letter code)
FLT_DATE	Network Manager	FLT_DATE	Date of flight
APT_ICAO	Network Manager	APT_ICAO	ICAO 4-letter airport designator
APT_NAME	PRU	APT_NAME	Airport name
STATE_NAME	PRU	STATE_NAME	Name of the State in which the airport is located
FLT_DEP_1	Network Manager	IFR departures - (NM)	Number of IFR departures from Network Manager
FLT_DEP_IFR_2	Airport Operator	IFR departures - (APT)	Number of IFR departures for which data was submitted
DLY_ATC_PRE_2	Airport Operator	ATC pre-departure delay (APT)	Minutes of delay reported as IATA Code 89
FLT_DEP_3	CODA/ Airlines	IFR departures - (AL)	Number of IFR departures for which data was submitted
DLY_ATC_PRE_3	CODA/ Airlines	ATC pre-departure delay (AL)	Minutes of delay reported as IATA Code 89

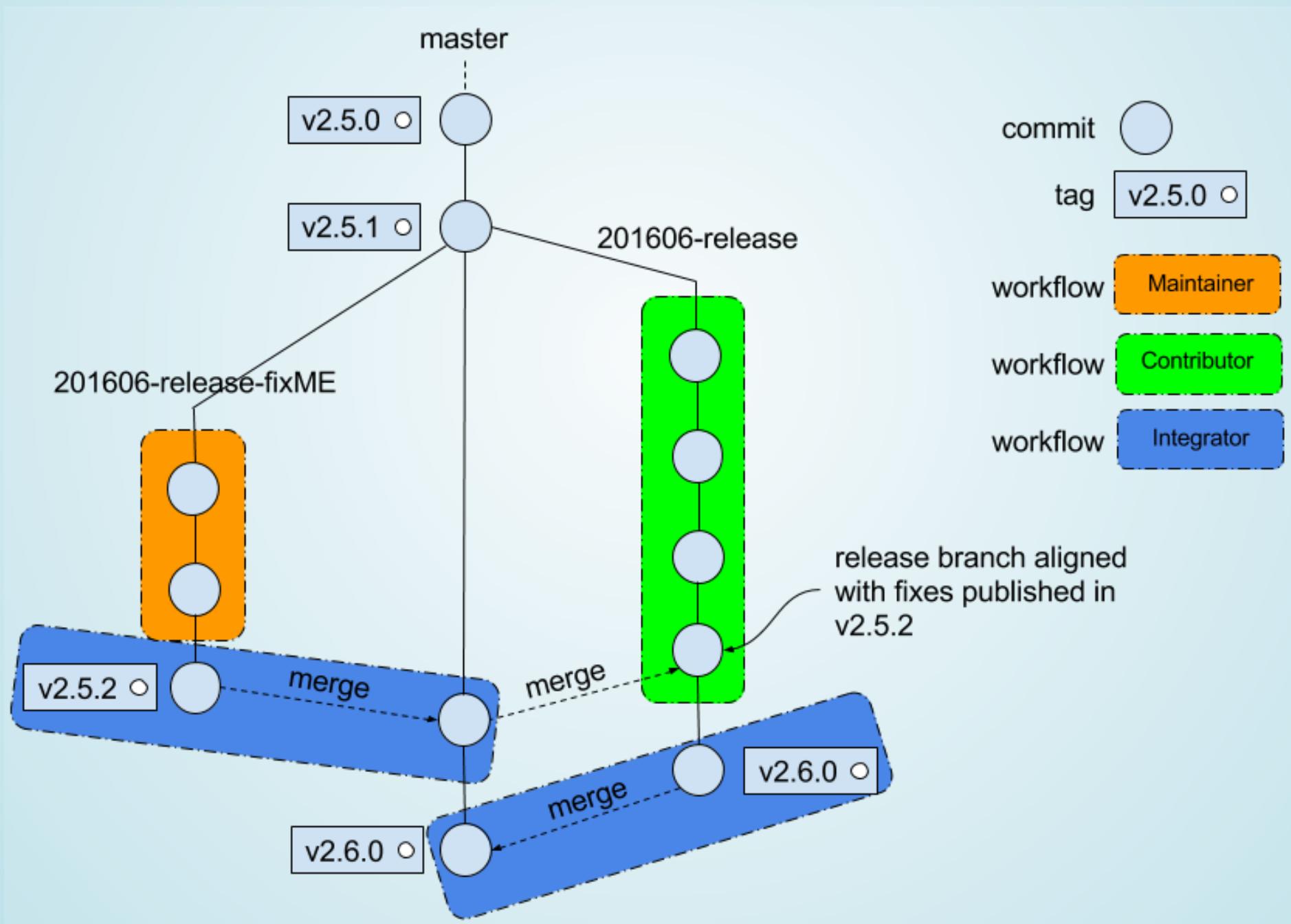
References

1. IATA. [Airport Handling Manual](#). IATA, 2012.

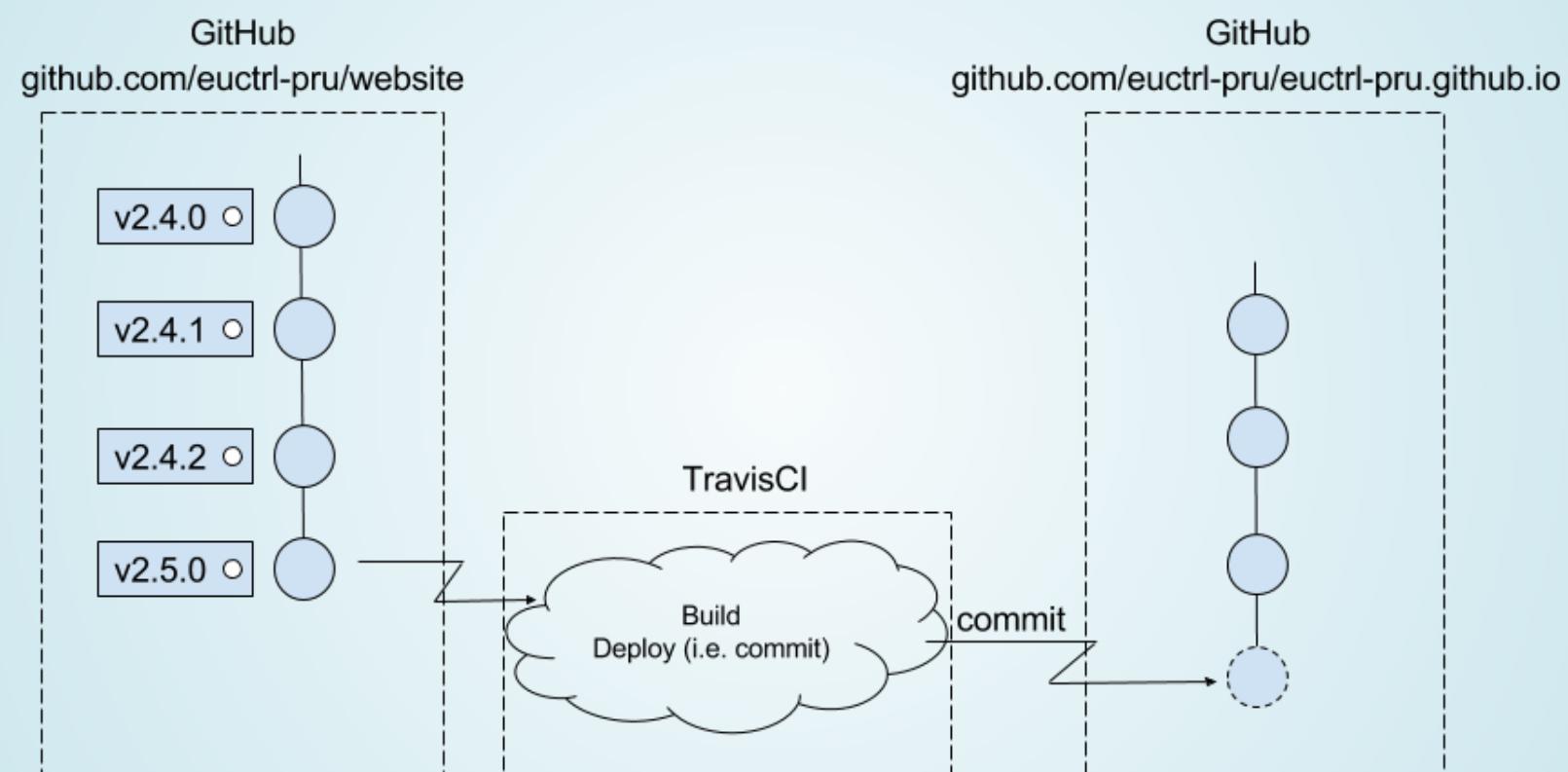
GENERATION

- from DB queries to website: scripts
- [Jekyll](#): MD -> HTML
- [Pandoc](#): MD -> PDF
- some from Rmarkdown/[[knitr](#)] in the near future

WORKFLOWS



TRIGGER



TRAVIS CI

Travis CI  Blog Status Help Enrico Spinielli 

Search all repositories 

[My Repositories](#) +

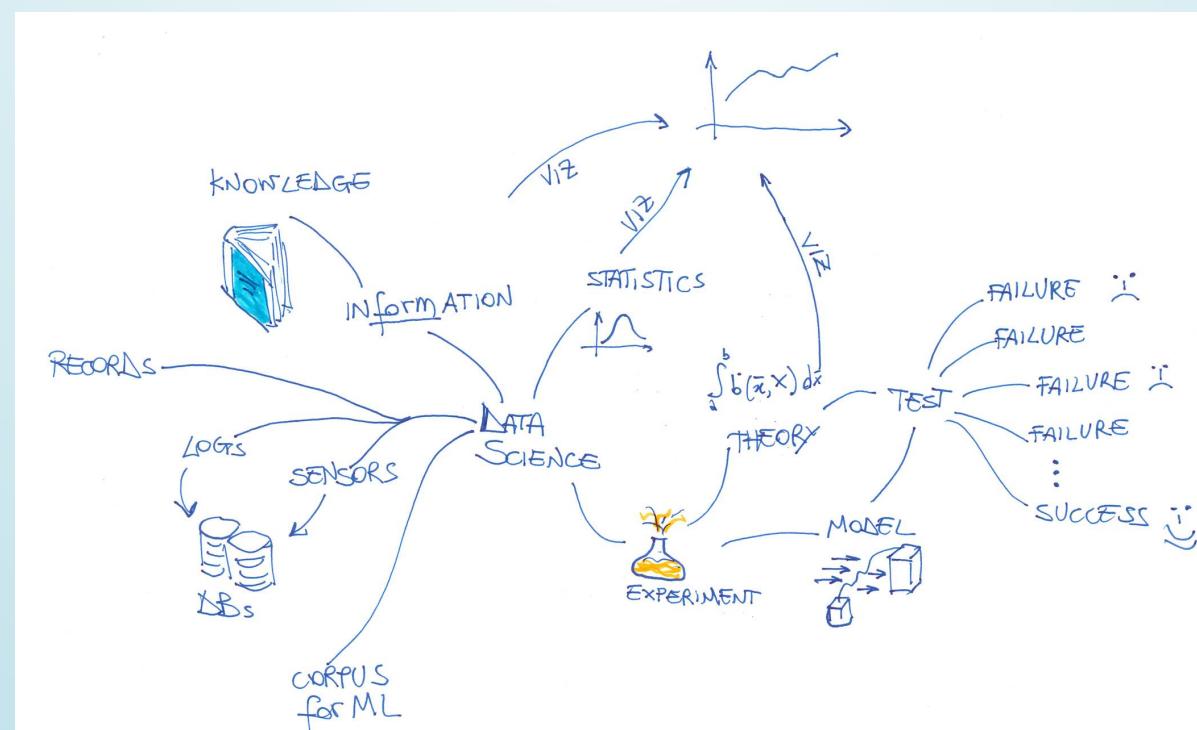
euctrl-pru / website  build passing

Current	Branches	Build History	Pull Requests	More options	☰
✓ euctrl-pru/website	# 124	✓ 201606-release added SEO plugin Enrico Spinielli	-o #124 passed 8bd172b	1 min 31 sec	9 days ago
		✓ 201606-release better tagging Enrico Spinielli	-o #123 passed 6478957	2 min 1 sec	11 days ago
		✓ master Updated ASMA dataset with correction to EDDP for Jan 2015 Enrico Spinielli	-o #122 passed 35f4faa	2 min 34 sec	11 days ago
		✓ 201605-release-fix07 Updated ASMA dataset with correction to EDDP for Jan 2015 Enrico Spinielli	-o #121 passed 35f4faa	1 min 59 sec	11 days ago
		✓ 201606-release added ICNS paper in Biblio Enrico Spinielli	-o #120 passed 17acb23	2 min 11 sec	12 days ago
		✓ master Merge branch 'master' of https://github.com/euctrl-pru/website into	-o #119 passed	2 min 45 sec	

TODO'S

DB

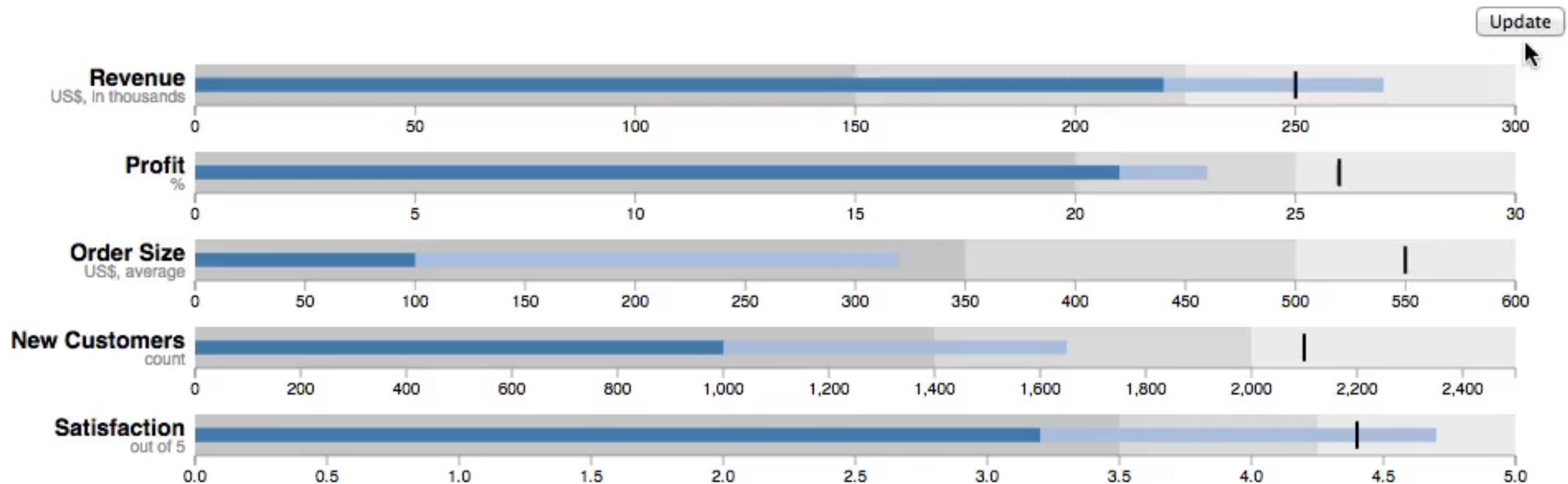
- new schema for production: PRUPROD
- use current ones for development (PRUDEV) and testing (PRUTEST)
- version control [PL]SQL code, i.e. which code was used to produce which indicators
- *version control* the DB used for prod: regulatory repository



DATA

- clarify dimensions
- improved the **Meta** part of it: definitions, methodology
- add more data and (web) API (see **ICAO iSTARS**)
- generate the spreadsheets if CSV files/API are not enough

VIZ



MORE VIZ

- more Studies/Articles w/ interactivity (see NYT, WP)
- more thinking of what is worth plotting
- more graphs in Graphs
- one year old experiment [www](#)
- a recent one w/ STATFOR [www](#)

CONCLUSIONS

WE WANT YOU!

- Share knowledge (or lack of)
- Learn from and know each other
- Discover internal and external datasets
- criticize & propose alternatives
- signal things you saw and would like to see implemented in our site
For example [NYT](#), Bloomberg ([1](#), [2](#)), [WP](#), [ProPublica](#), [The Guardian](#),
[Financial Times](#) ... have fantastic infographics





WE HEAR YOU!

- emails with questions, proposals are a good start
- you are always welcome to come and chat (but bring your coffee)
- present at the next *Show & Tell*

The image features a large, bold, purple word "thanks" centered in the foreground. Behind it, numerous other words representing "thank you" in various languages are scattered across the page. These include "Ahó" (pink), "Obrigadu" (purple), "Juspajaraña" (yellow), "Fakaaue" (green), "Néáeše GV" (light green), "Quyanaq" (pink), "Chokran" (red), "Na gode" (red), "Mersi" (green), "Mil" (blue), "Ga" (blue), "Ura" (red), "Sipas" (pink), "Teskükür edirəm" (green), "Mälö" (brown), "Tänan" (brown), "Tak" (blue), and "Yin acar" (orange). The background is white, and the overall composition is a collage of international gratitude expressions.

Takk

REFERENCES AND INSPIRATIONS

TOOLS

- Git
- GitHub
- TravisCI
- D3.js
- Vega
- Google Charts (w/ caveats)
- R: ggplot2, dplyr, R Studio

SOCIAL

- D3.js: [Mike Bostock](#), [Ian Johnson](#), [Elijah Meeks](#), [Nadieh Bremer](#), [Susie Lu](#), [Christophe Viau](#)
- Viz gurus: [Alberto Cairo](#), [Stephen Few](#), [Edward Tufte](#), [Enrico Bertini](#), [Maarten Lambrechts](#), [Jonathan Corum](#), [Jeffrey Heer](#) ...
- Twitter: big inspiration from `#d3js #dataviz`
- Statistics: [Hadley Wickham](#)
- Conferences: [Eyeo Festival](#), [OpenVis Conference](#)
- Awards: [Malofiej](#), [Data Journalism Awards](#)
- Newspapers: [NYT](#), [The Guardian](#), [Bloomberg](#), [The Washington Post](#), [Pro Publica](#), [National Public Radio](#) ...

BOOKS

Yes, you still have to study!

- Tufte, Edward
- Cairo, Alberto
- Few, Stephen

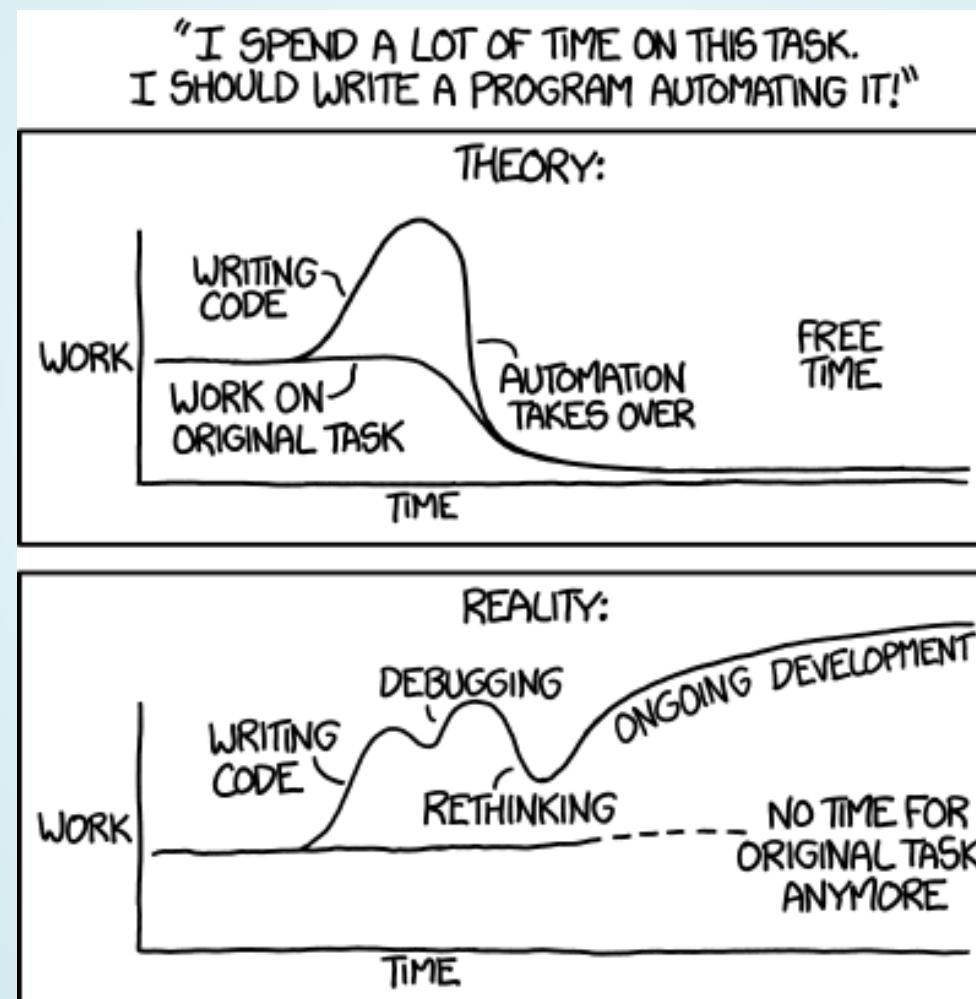
CREDIT WHERE IT IS DUE

- PRU animated logo inspired by UW Interactive Data Lab
- Thank You cloud generated w/ Jason Davies' Word Cloud generator
- Search Engine functionality from DuckDuckGo API
- Flight Flows inspired by [Global Flow of People](#) and made possible by adapting [Vienna Institute of Demography](#) shared [source code](#)
- STATFOR's experiment inspired by Amanda Cox's [NYT](#) article via Vega Editor example

TRIVIA

AUTOMATION 1

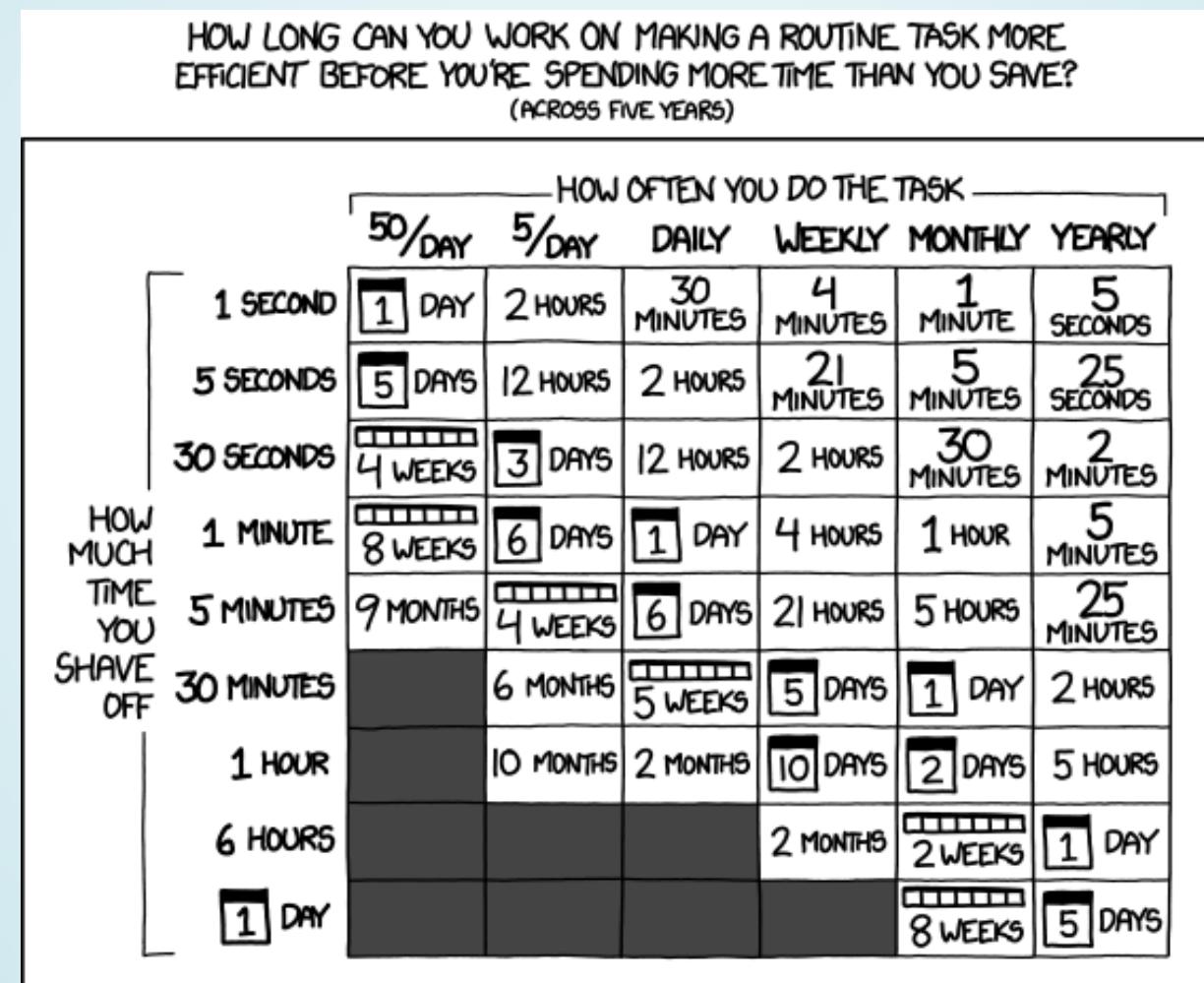
xkcd 1319 and explanation



Title text: 'Automating' comes from the roots 'auto-' meaning 'self-', and 'mating', meaning 'screwing'.

AUTOMATION 2

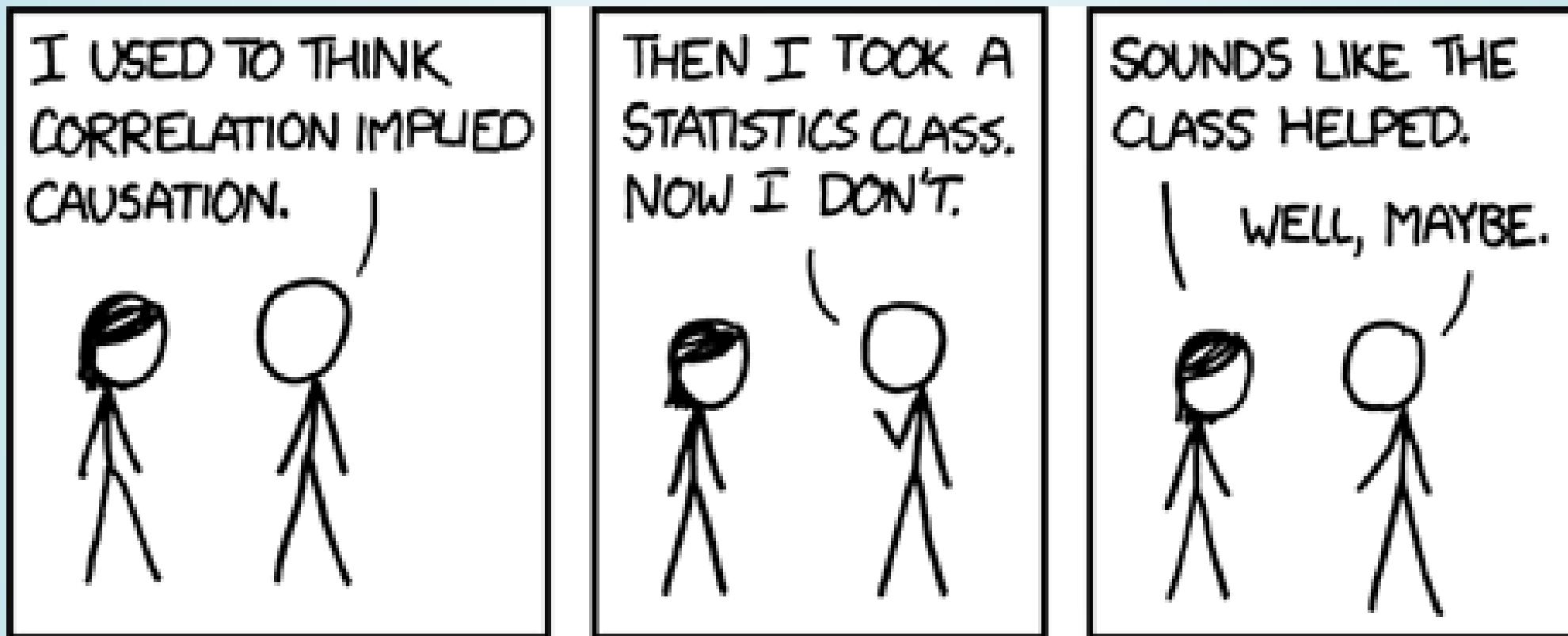
xkdc 1205 and explanation



Title text: Don't forget the time you spend finding the chart to look up what you save. And the time spent reading this reminder about the time spent. And the time trying to figure out if either of those actually make sense. Remember, every second counts toward your life total, including these right now.

CORRELATION

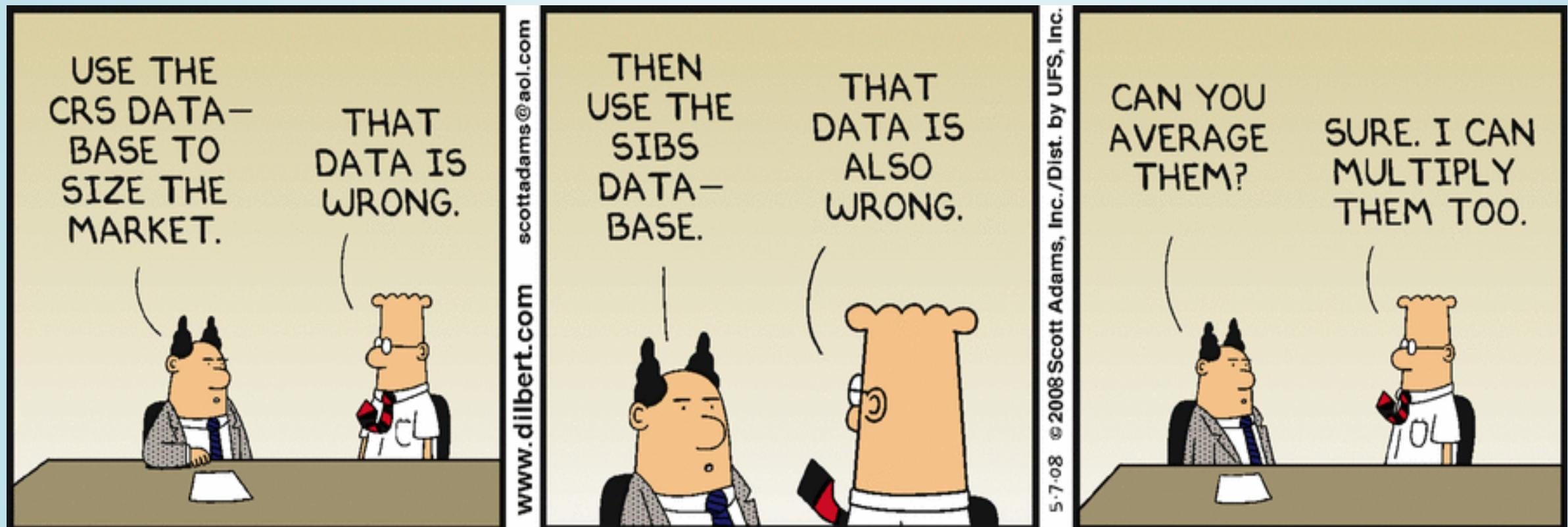
xkcd 1205 and explanation



Title text: Correlation doesn't imply causation, but it does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there'.

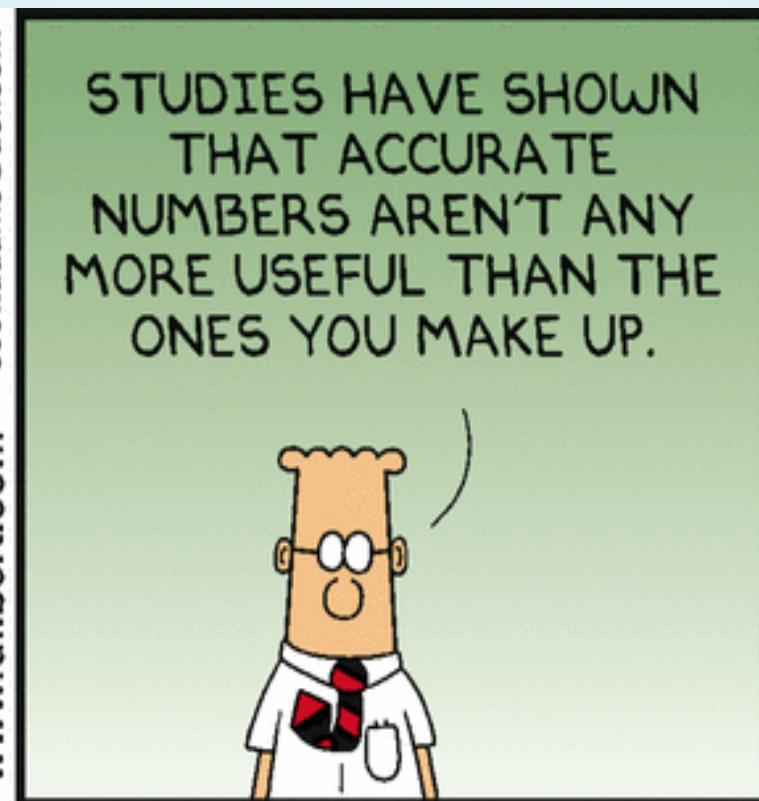
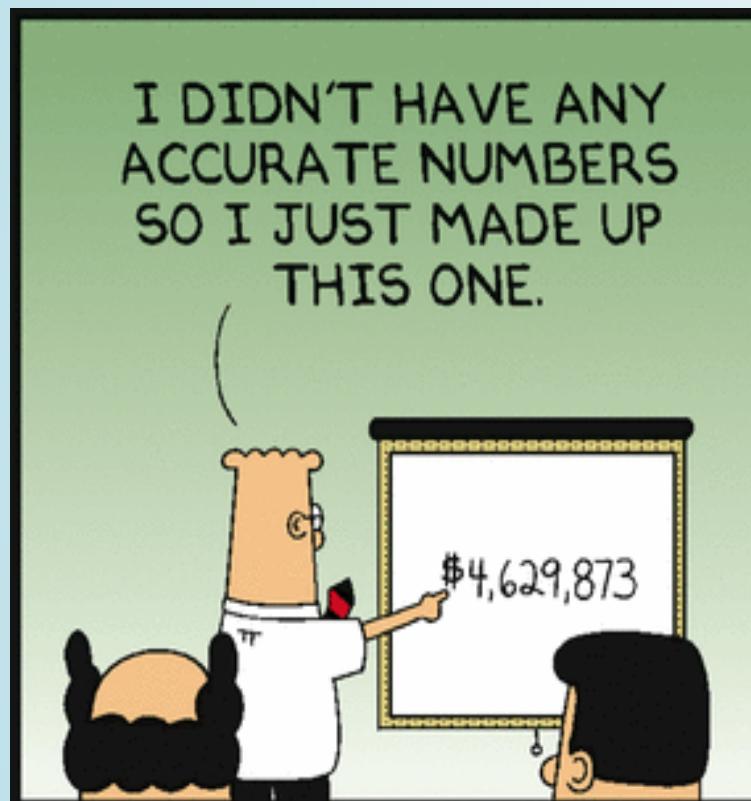
DATA ACCURACY 1

Dilbert 2008-05-07



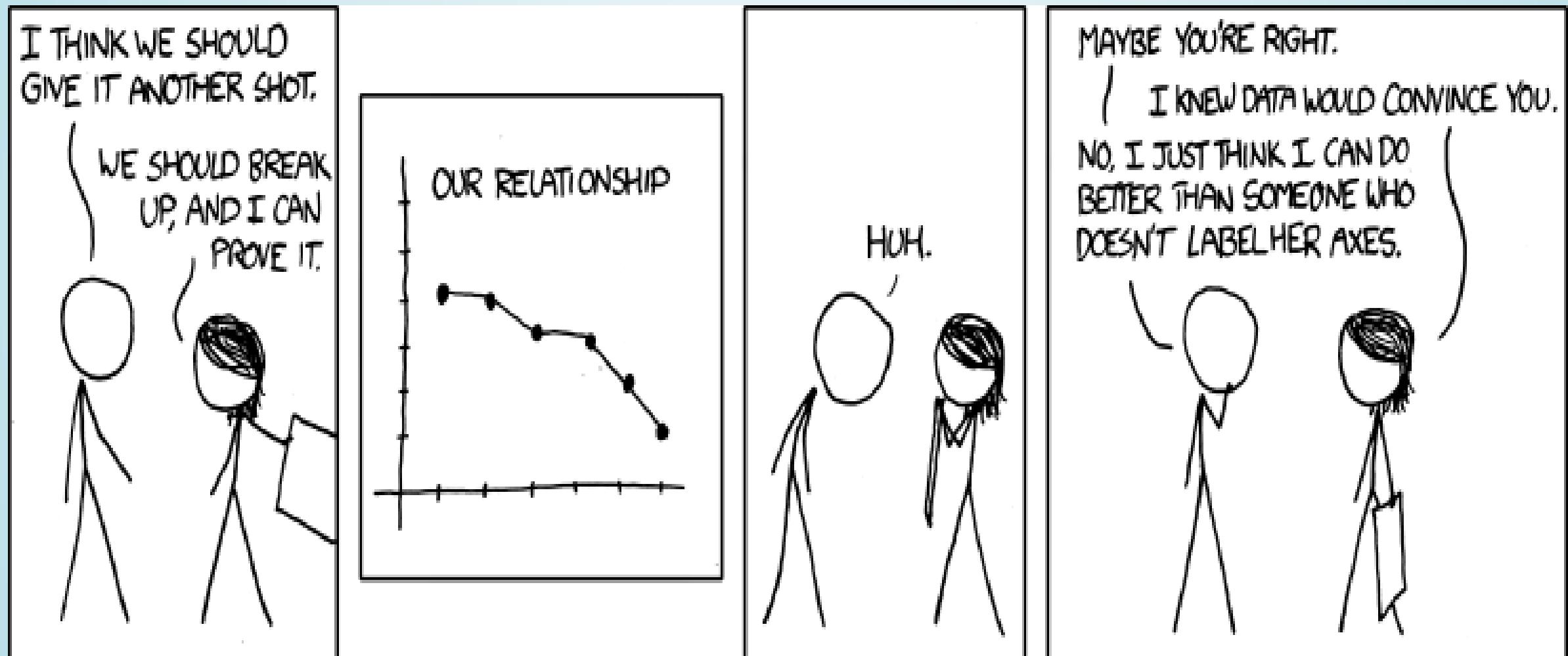
DATA ACCURACY 2

Dilbert 2008-05-08



CONVINCING

xkdc 833 and explanation



Title text: Don't forget the time you spend finding the chart to look up what you save. And the time spent reading this reminder about the time spent. And the time trying to figure out if either of those actually make sense. Remember, every second counts toward your life total, including these right now.

BIG DATA

