Announcement 1: Melbourne Datathon 2017

- 13 April - May 5 (3 weeks)
- 2 competitions
  - Insights
  - Kaggle
- New dataset

- Free event, free pizza, free drinks
- Cash prizes ($5k)

http://www.datasciencemelbourne.com/datathon/

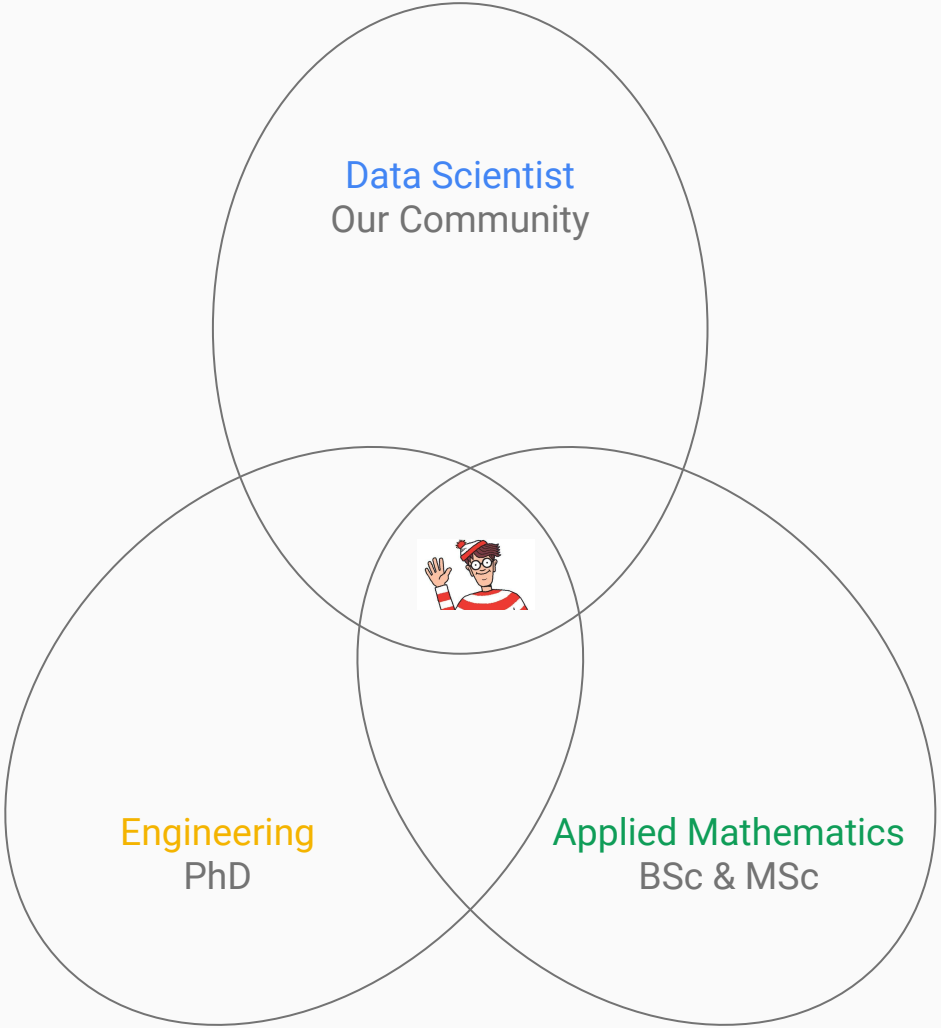
Announcement 2: Melbourne Data Science Week 2017

- 29 May - 2 June
- Tutorials
- Networking
- Conference ($70 for students)

http://www.datasciencemelbourne.com/medascin2017/

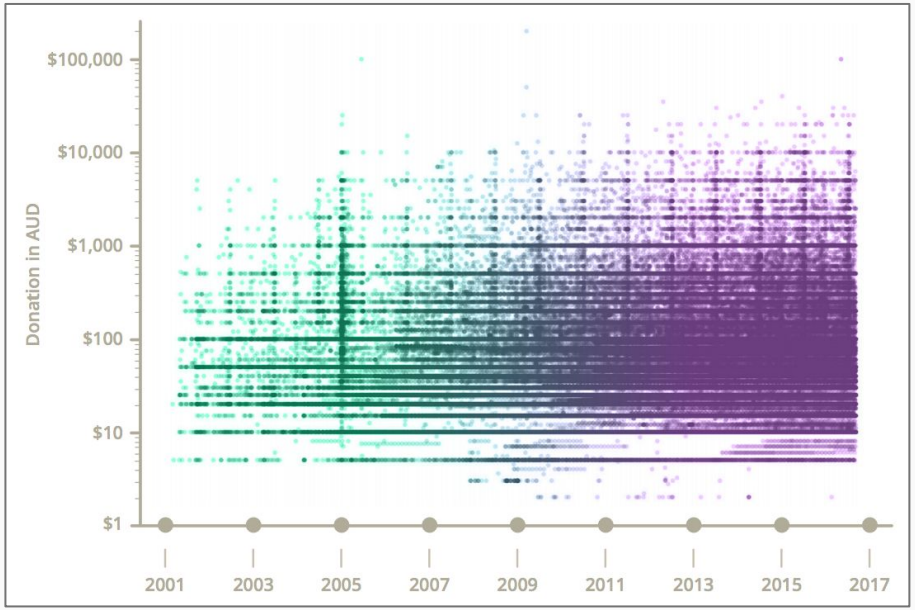
From Student to “Scientist”

Joost van der Linden


Data Scientist
Our Community
Engineering
PhD
Applied Mathematics
BSc & MSc

Data Scientist

Our Community



https://www.givenow.com.au/stats

Data Scientist

Our Community

CLASSIE

Classification of Social Sector Initiatives & Entities

SmartyGrants

Software & data science for revolutionary grantmakers

What is the difference?

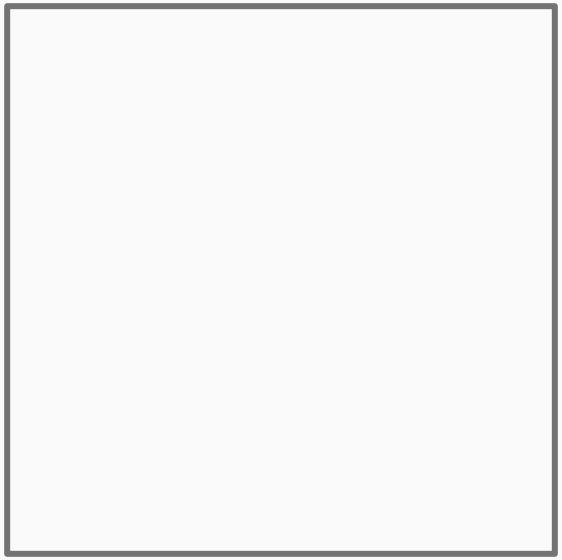
Communication

Communication

Harvard Business Review, January 2017

From ambiguity to clarity.

Solve for x



Solve for A and B and don't forget about C. Build platform D to inform E about F. Show G in terms of H and I and sort by J. Follow rule K to inform L about change in M, correlating N with O while accounting for P. Study Q, forget about R. Test S compared to the T. Change U because exploring V said W. Definitely assume X. Extract Y.

Solve for A and B and don't forget about C. Build platform D to inform E about F. Show G in terms of H and I and sort by J. Follow rule K to inform L about change in M, correlating N with O while accounting for P. Study Q, forget about R. Test S compared to the T. Change U because exploring V said W. Definitely assume X. Extract Y.

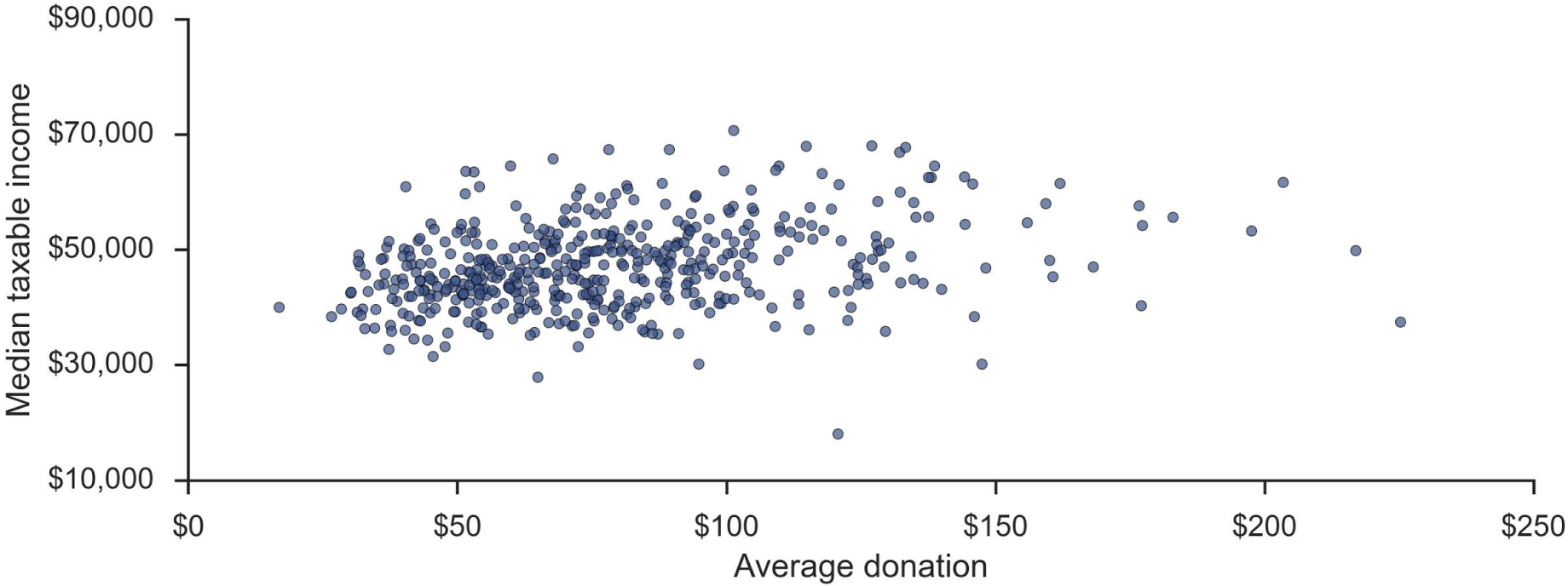
Communicating complexity

Communicating complexity

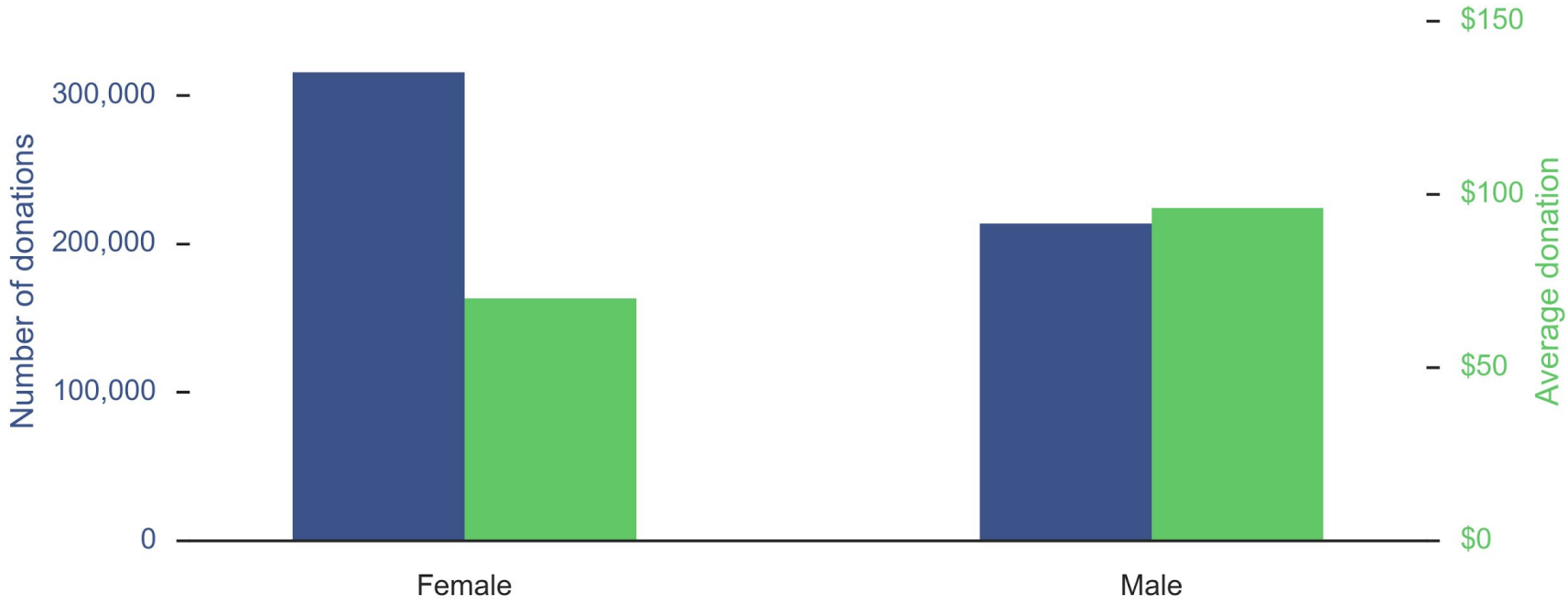
As a student, your audience is smarter than you.

As a data scientist, you are the educator.

Communicating complexity



Communicating complexity



Real data

X Y Z

User interfaces

- Human complexity in the input data
- Human complexity in processing the data
- Human complexity in showing the data

Actionable insights

Actionable insights

“It is easy to get caught up in paralysis by analysis, where we go down these rabbit holes of interesting factoids that can never be translated into action”

- Andrew Means

Over-engineering

Over-engineering

The deep dive

- Phrase modeling
- Latent Dirichlet Topic Modeling
- Word vector models with Word2Vec
- Visualising Word2Vec with t-SNE

Great tutorial: Modern NLP in Python

Over-engineering

The deep dive

- Phrase modeling
- Latent Dirichlet Topic Modeling
- Word vector models with Word2Vec
- Visualising Word2Vec with t-SNE

The shallow dive

- Word frequencies
- n-Grams
- Wordclouds
- Reading the text itself...

Great tutorial: Modern NLP in Python

Grey areas

Grey areas

Grantmaker: “I want to see where my money is going”

Grey areas

Grantmaker: “I want to see where my money is going”

Our Community: “Let’s plot $ amount by subject, as a bar chart”

Grey areas

Grantmaker:	“I want to see where my money is going”
Our Community:	“Let’s plot $ amount by subject, as a bar chart”
Grant applicant:	“I can’t divide my funds like that”

Grey areas

Grantmaker: “I want to see where my money is going”

Our Community: “Let’s plot $ amount by subject, as a bar chart”

Grant applicant: “I can’t divide my funds like that”

Ever-lasting debate between usability and data-quality.

Take-homes

Take-homes

Newsletters

- Data Elixer
- Data Science Weekly
- Data Machina
- KD Nuggets
- Roundup
- O'Reilly

Take-homes

Portfolio

1. Learn Git
2. Fork a template (fork my website)
3. Put your stuff up

Be smart about the projects you get involved in.

Take-homes

Portfolio

1. Learn Git
2. For someone else's website
3. Put your stuff up

Be smart about the projects you get involved in.

Getting a job

1. Network, Meetup, Datathon
2. And/or let your work speak for you

Take-homes

“The confusing messages about the data science career”

- u/sasquatch007 on r/datascience

For the things we have to learn before we can do them, we learn by doing them.

- Aristotle