

# Pokemon, Big Data and Everything

# Pokemon, BigData and Everything

**Why econometricians aren't *that* useful.**

# **who is this guy?**

# My Story

I was about to graduate econometrics 4 years.

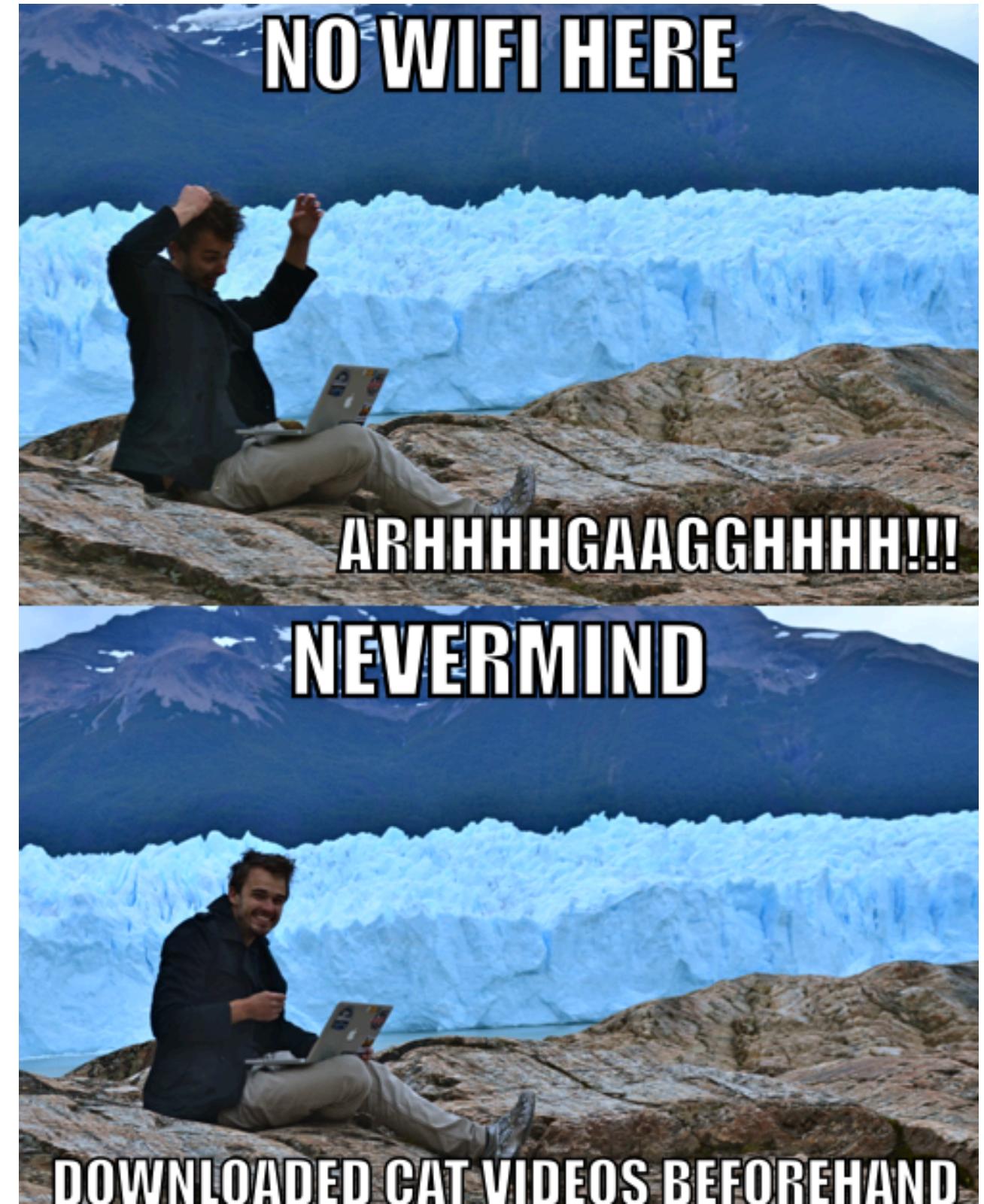
- I worked for a business intelligence company.
- Wore a suit to work everyday and I clicked on a lot of buttons to make powerpoint slides. Such was the life of the post-student enterprise lifestyle.

Then my boss asked me for a rather meaningless task ... this sort of changed my life from guy-in-suit to an internet meme.

# THE WIFI SUCKS HERE



# WANT MONEY BACK



**"As much as I love cat videos, who actually downloads them and maintains a stash?"**



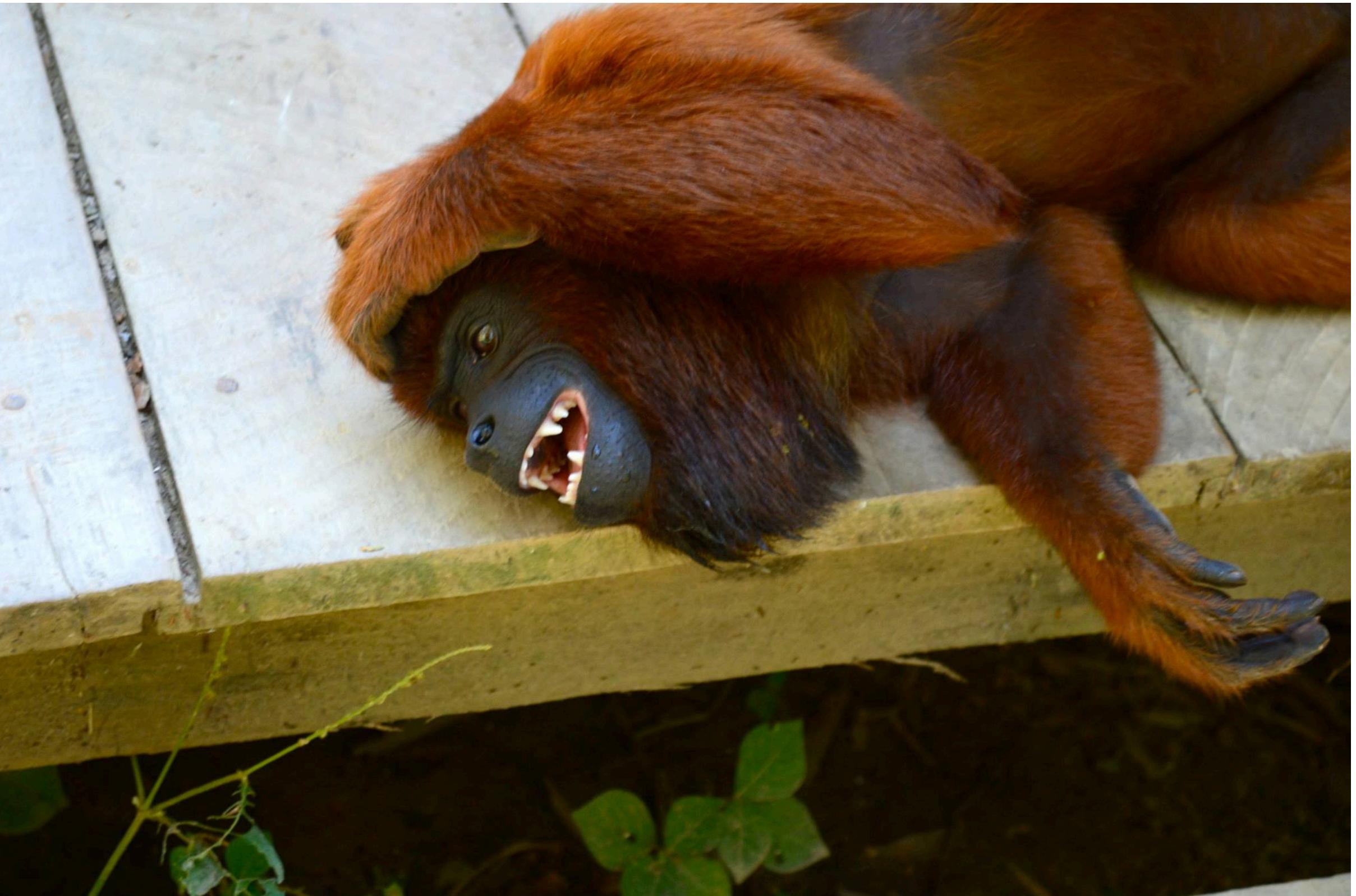
**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**



**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**



**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**



**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**



**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**



**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**



**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**



**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**



**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**

# **Fast Forward a while**

# I'm also this guy.

Europe



Mango Business Solutions Ltd  
UK



Mango is a leading data science and data analytics company offering customized solutions, training, consultancy, application development and support.

[+](#) Click for Description



Vincent Warmerdam  
Amsterdam, Netherlands



Vincent D. Warmerdam is a data scientist at GoDataDriven, the leading big data science consultancy firm in the Netherlands.

[+](#) Click for Biography



Vincent Guyader  
France



Vincent Guyader is a french statistician and President of ThinkR, a company which helps businesses migrate to R. He teaches R in finance, insurance, marketing, and life sciences.

I'm also this guy.



# I'm also this guy.

**Meetup**

## The Amsterdam Applied Machine Learning Meetup Group

Home Members Sponsors Photos Pages Discussions More      Group tools My profile

Amsterdam, Netherlands  
Founded Oct 10, 2013

About us...  
+ Invite friends

machine teachers 994  
Group reviews 7  
Past Meetups 9  
Our calendar

Organizers:  
 **Friso van Vollenhoven**, Gijs Molenaar, Vincent Damian Warmerdam

Copy this Meetup

### Streaming Recommendations and Model Factories

October 21 · 6:00 PM  
GoDataDriven

Rather overdue, but here it is: the next Applied Machine Learning  Meetup! This e... [See all](#)



114 went

 **Vincent Damian Warmerdam**  
Co-Organizer, Event Host

 **Gijs Molenaar**  
Co-Organizer, Event Host

Good to see you

 **domagoj fizulic**

Good to see you

How was the Meetup?   
avg: 

Ask a question, share something, or leave a comment...

Post

**Vincent D. Warmerdam - @fishnets88 - koaning.io - GoDataDriven**

# I'm also this guy.

Eventbrite |  Search for events | BROWSE EVENTS | HELP | VINCENT | CREATE EVENT

⌚ This event has ended

## Open R Session

Vincent D. Warmerdam  
Saturday, 28 May 2016 from 10:00 to 17:00 (CEST)  
Amsterdam, Netherlands



**Ticket Information**

TYPE	REMAINING	END	QUANTITY
RSVP	1 Ticket	Ended	Free
			N/A

**Who's Going**

Oops! We're having trouble connecting to Facebook. Please [try again](#).

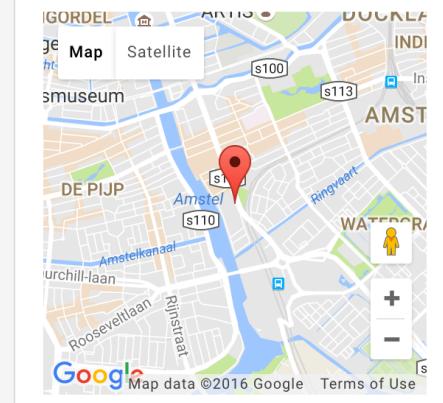
**Share Open R Session**

[!\[\]\(8c28abcc812daa7358c8fce8c8d0920d\_img.jpg\) Share](#) [!\[\]\(9ce535ae3f041d7c0eb1883b286082dc\_img.jpg\) Tweet](#) [!\[\]\(c2a4ef8aba0fe183743185a538adffc6\_img.jpg\) Like](#) Be the first of your friends to like this.

**Event Details**

**Open R Session**

**When & Where**



Xebia GoDataDriven  
Wibautstraat 200/202  
1091 GS Amsterdam  
Netherlands

Saturday, 28 May 2016 from 10:00 to 17:00 (CEST)

[!\[\]\(6e2fcb89970d41a52ce6628b942539cc\_img.jpg\) Add to my calendar](#)

# I'm also this guy.



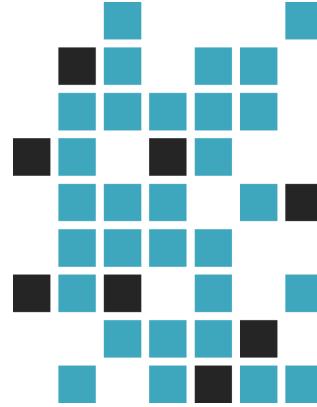
python, javascript, R, dplyr, ggplot, ditto, scala, mongo,  
html5, bootstrap, git, sublime, d3, leaflet, sawk, pandas,  
feebas, numpy, scikit, nltk, crebase, jupyter, onyx,  
lodash, mesos, docker, django, flask, neo4j, vulpix,  
selenium, node-webkit, hadoop, hoopa, impala, spark,  
azurill, ansible, hadoop, mapreduce.

# I'm also this guy.



python, javascript, R, dplyr, ggplot, ditto, scala, mongo,  
html5, bootstrap, git, sublime, d3, leaflet, sawk, pandas,  
feebas, numpy, scikit, nltk, crebase, juypter, onyx,  
lodash, mesos, docker, django, flask, neo4j, vulpix,  
selenium, node-webkit, hadoop, hoopa, impala, spark,  
azurill, ansible, hadoop, mapreduce.

Can you recognize which ones are pokemon?



# GoDataDriven

Vincent solves data problems, ask him anything. He has a blog about less obvious aspects in the world of data science over at [koaning.io](https://koaning.io) and he's known for giving free lectures in data science around Europe. He's especially keen to give one in Lisbon, Ljubljana, Split or Belfast. Feel free to notify Vincent if you've got a venue in any of these places or other awesome place to visit. He'll actually show up.

# **Long Story Short**

**my name is vincent  
i do open source data  
ask me anything**

# Today

- I'll talk to you guys about what I've been up to.
- I have mixed feelings about econometrics.
- I'm sure happy I thought myself programming.
- I'm very happy I don't wear a suit to work.
- Explain the difference between my definition of a data scientist and business analyst
- Talk about how I solve problems.

# Difference between data analyst/ data scientist.

With two examples I'll try to explain why we **do not** consider econometricians to be unicorns.

- pokemon
- world of warcraft
- if(time) pokemon part 2 (bayesians!)

# Conclusions

It's fashion to call yourself a data scientist. This hype comes from somewhere and this somewhere isn't a click-tool. I'd like to avoid any confusion so I would propose to stick to this sniff test.

```
if(SQL + Clicky Clicky) -> Data Analyst  
if(code + algorithms) -> Data Scientist
```

The alternative is to put pokemon names on my resume.

# Conclusions

In a lot of ways the world has opened up. It's getting very expensive not to use open source software.

In this new open jungle, you'll become less and less relevant if you cannot communicate with machines. It's the only way to automate the boring stuff.

Even microsoft is trying to remain relevant by switching to open source. Think about that.

# Conclusions

People treat econometricians like they are the magical unicorns of business. The ability to interpret a linear regression does not turn you into Gandalf the White. The new school computer geeks will eat that alive.

Calculus, probability theory and linear algebra are everything you need on a theoretical basis. Creativity is much more important.

# Conclusions

It also seems like there are too many people worrying about making moarmoney[tm]. Maybe we should stop making powerpoints and word files for McWaterHouse Deloitte Group and actually start making solutions to problems that matter.

# Conclusions

We enjoy artificial intelligence because we suffer natural stupidity. This doesn't mean that you cannot apply natural intelligence before you attempt to apply fancy models.

By the way, there's also a thing such as artificial stupidity.

# Conclusions

If you want to learn about this stuff; life has never been easier.

Check udacity, coursera, meetups and conferences. The open jungle is a wild place, but people are nice and don't mind teaching anyone.

Especially if you're a women you should feel especially welcome. The gender imbalance in the field is suspicious enough to call it a problem.

# Conclusions

If I'm going to send you home with anything though ...

**"Never let your school get in the  
way of your education."**

**Mark Twain**