

# **PhD Proposal with Python**

## **o como convencer a ingenieros que hagan ciencia**

Ing. Cabral, Juan B.



Universidad Nacional del Sur

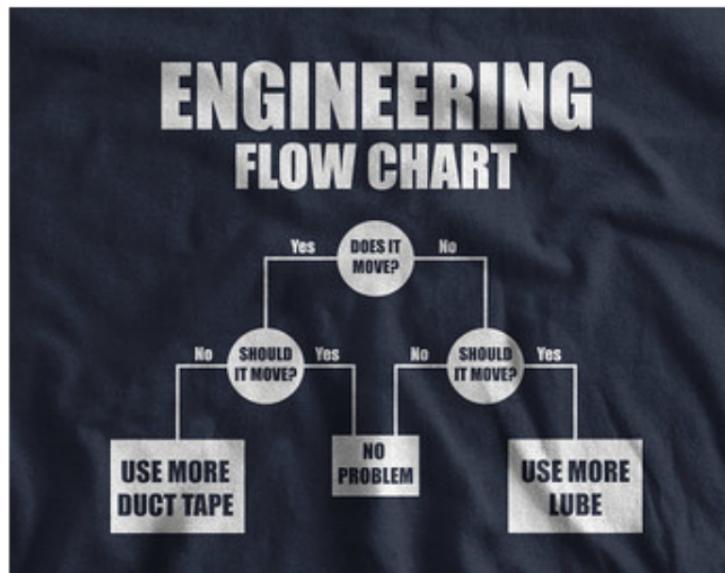
SciPyCon Argentina 2014

10/2014 - Bahía Blanca - Argentina

# About Me

## About Me

- Software engineer.
- Data scientist.

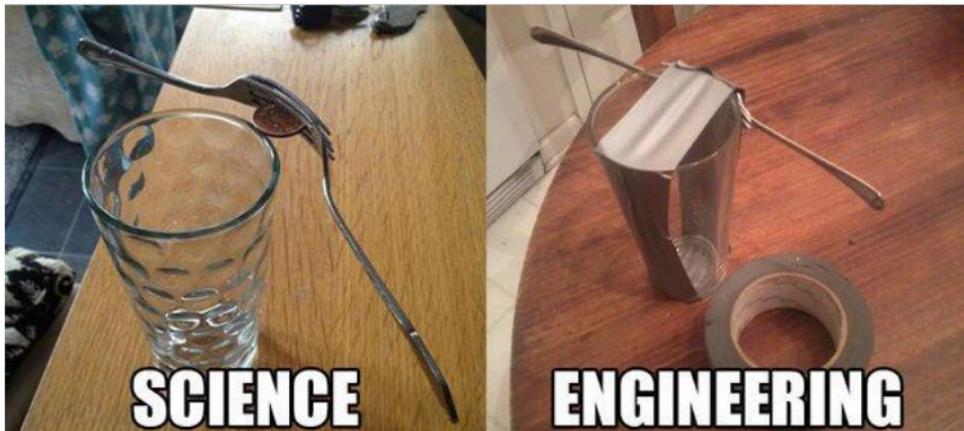


# Agenda

- De ingeniero a científico.
- Encontrando un problema interesante.
- Que herramientas me sirvieron para justificar el plan.
- Conclusiones.



# De ingeniero a científico - Ciencia vs Ingeniería



## La ciencia y la ingeniería (investigación vs diseño) [\[editar\]](#)

La ciencia, investiga, le interesa saber, su producto son los conocimientos.

La ingeniería por su lado, aplica todos aquellos conocimientos que son el resultado de la investigación. Le interesa el conocimiento de la ciencia en la medida en que lo pueda aplicar; el producto son las obras y los aparatos físicos que crea.<sup>1</sup>

# De ingeniero a científico - Crecimiento de datos

 **65 billion**

Location-tagged payments  
made in the U.S. annually

 **154 billion**

E-mails sent per day

 **87%**

U.S. adults whose location is  
known via their mobile phone

## Digital Information Created Each Year, Globally

2,000 BILLION GIGABYTES

1,800

1,600

1,400

1,200

1,000

800

600

400

200

0

2005 2006 2007 2008 2009 2010 2011

**2,000%**

Expected increase in  
global data by 2020



**Megabytes**

Video and photos stored  
by Facebook, per user

**75%**

Percentage of all digital  
data created by consumers

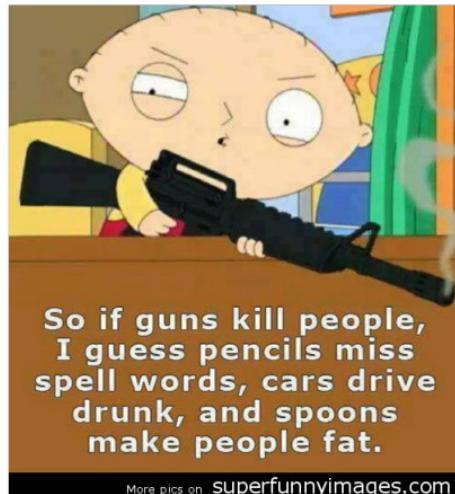
Sources: IDC, Radicati Group, Facebook, TR research, Pew Internet

# De ingeniero a científico - Las herramientas actuales



## De ingeniero a científico - Algunas conclusiones previas

- El crecimiento de datos se lo esta llamando **revolución**
- El manejo de herramientas para grandes volúmenes de datos **NO** es trivial
- El diseño de conocimiento nuevo para el manejo de las técnicas y herramientas de análisis se ha transformado en una especialidad en si mismo.
- Los ingenieros estamos para ayudar.



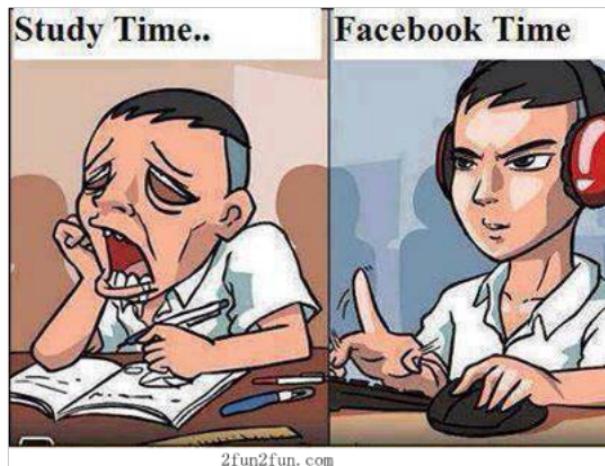
# Agenda

- De ingeniero a científico.
- **Encontrando un problema interesante.**
- Que herramientas me sirvieron para justificar el plan.
- Conclusiones.



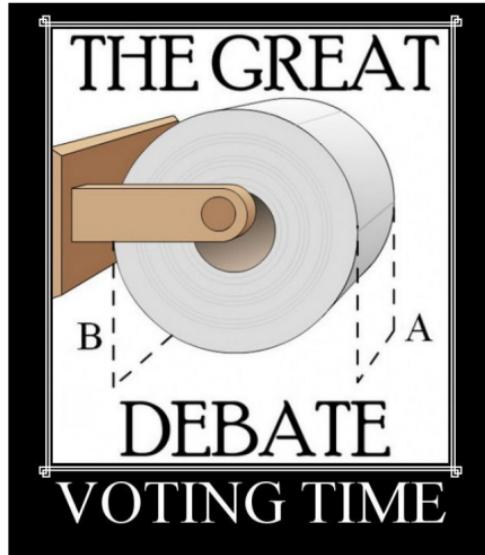
# Un problema interesante - Conocimientos Previos

- Trabajo en un grupo de investigación sobre bioinformática en la UTN.
- Me especialicé en estadística orientada a la investigación.
- Trabaje diseñando modelos (bastante artesanales) para análisis de información en una empresa de análisis de medios masivos.
- Trabaje 3 años diseñando soluciones basadas en soporte de Business Intelligence



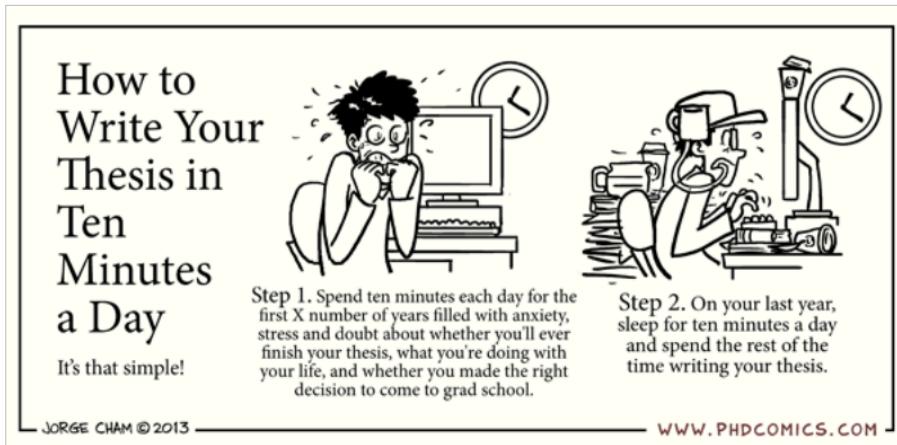
## Un problema interesante - Alternativas

- Hago un doctorado?
- Lo hago en disciplinas que son de interés a la ingeniería?
- Busco una necesidad de terceros?

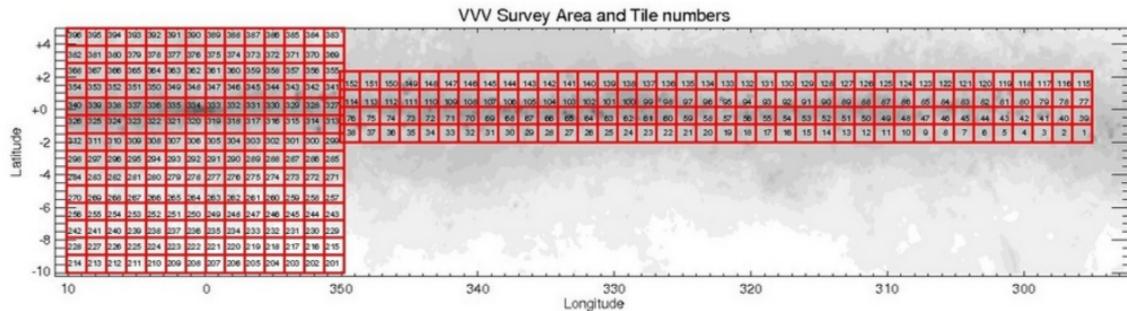


# Un problema interesante - Eligiendo el DP

- En la ScipyCon Argentina 2013 Sebastián Gurovich (IATE-OAC) hablo de un problemita de datos de un relevamiento llamado el VVV
- Seguimos charlando post conferencia en Córdoba.
- Elegí la universidad donde realizar la titulación.
- Escribimos un plan "draft" y buscamos un codirector.



# Un problema interesante - El problema



# Agenda

- De ingeniero a científico.
- Encontrando un problema interesante.
- **Que herramientas me sirvieron para justificar el plan.**
- Conclusiones.



# Herramientas



<http://www.astropython.org/resources>

# Conclusiones

- El advenimiento de datos ya cambió la forma de hacer ciencia en la ultima década.
- No me parece (IMHO) que se este capacitando a ingenieros para asistir a tareas científicas (al margen de casi no hay ingenieros)
- El análisis de datos ya esta generando mas roles que antes no existían como el astro-estadístico.
- No todas las currículas de ingeniería te dejan preparado para investigar (en mi casi fue un esfuerzo personal)



# ¿Preguntas?

- Charla:

- Contactos:

- [jbcabral.com](http://jbcabral.com)
- Juan B Cabral <[jbc.develop@gmail.com](mailto:jbc.develop@gmail.com)>

