Thinking with data

Junio 2017

Introduction

Extractado del libro Thinking with data[1].

[1] MacNamara Fallacy: measurement is not understanding.

Project scope

There are four parts to a project scope[2]:

[2] Variant: context, objectives, how does it work?, next steps

1. **the context**,
2. **the needs** trying to meet[3]. It's important to understand how meeting those needs will help the organization achieve its larger goals,
3. **the vision** of what success might look like. It could consist of a **mockup** describing the intended results, or a **sketch of the argument** that we're going to make.
4. **the outcome**, in terms of how the organization will adopt the results and how its effects will be measured down the line.

[3] Peter Hintjens: los proyectos deben partir de los problemas y no de las ideas. Los proyectos de data science resuelven la falta de conocimiento, arrojar luz, y otras disciplinas resuelven otras necesidades. En nuestro caso las necesidades del proyecto han de estar encuadradas en las necesidades de nuestro equipo y de la organización.

Arguments

Data consists of **observations** about the world (records in a database, notes in a logbook, images in a hard drive). **Arguments** are what make knowledge out of observations. Only in mathematics is it possible to demonstrate something beyond all doubt[4]. An argument moves from statements that the audience already believes to statements they do not yet believe.

Arguments are built around claims.

There is no single statistical tool that is sufficient to create confidence in either a causal relationship or the knowledge that a pattern observed should continue to hold in the future..

Evidence is the introduction of facts into an argument. Our evidence is rarely raw data[5].

[4] Las pruebas no demuestran que el software funcionará siempre más allá de toda duda.

[5] Raw data needs to be transformed into something else, something more compact, before it can be part of an argument.

Patterns of Reasoning

A **dispute of fact** turns on what is true, or on what has occurred. There are two stock issues for disputes of fact:

- What is a reasonable truth condition?
- Is that thruth condition satisfied?

Disputes of definition occur when there is a particular way we want to label something, and we expect that that label will be contested. There are two activities involving definitions that can happen in an argument. The first is making a case that a general, imprecise term fits a particular example[6]. The second activity is that we have made an existing term more precise[7]. There are three stock issues with this kind of disputes:

- Useful. Does this definition make a meaningful distinction?
- Consistent. How well does this definition fit with prior ideas?
- Best. What, if any, are the reasonable alternatives, and why is this one better?

When we are concerned with judging something, the dispute is one of **value**[8]. For disputes of value, our stock issues are:

- How do our **goals** determine which values are the most important for this argument?
- Has the value been properly applied in this situation?

Disputes of **policy** occur whenever we want to answer the question, Is this the right course of action? or Is this the right way of doing things?. The four stock issues of disputes of policy are:

- Ill. Is there a problem?
- Blame. Where is credit or blame due?
- Cure. Will the proposal solve it?
- Cost. Will it be better on balance?

[6] Por ejemplo si mostramos unos gráficos y concluimos que una empresa está creciendo. ¿Qué significa creciendo?

[7] Por ejemplo si afirmamos que hemos desarrollado reglas para determinar qué lenguajes de programación son los más populares o qué usuarios son los más influyentes en Twitter. Los terminos popular e influyente son imprecisos pero al definir esas reglas los estamos clarificando.

[8] Alguno es bueno o malo. Algo es mejor o peor.