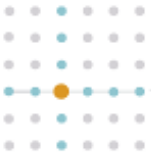
5 lessons I have learned at

Analyx®

Strategic Predictive Customer Insights

Who I am?

- Data scientist at Analyx since 2013
- PhD in Business and Quantitative Methods (UC3M, Spain)
- Industrial Engineer
- useR since 2005, and active member of the R community.
- Self promotion: www.adolfoalvarez.cl, @adolfoalvarez



Analyx®

Strategic Predictive Customer Insights

Who we are?

What do we do?

- Marketing Analytics
- New products simulations
- Media Mix Modelling
- Sales modelling
- Churn prediction

How do we do it?

This leads to…

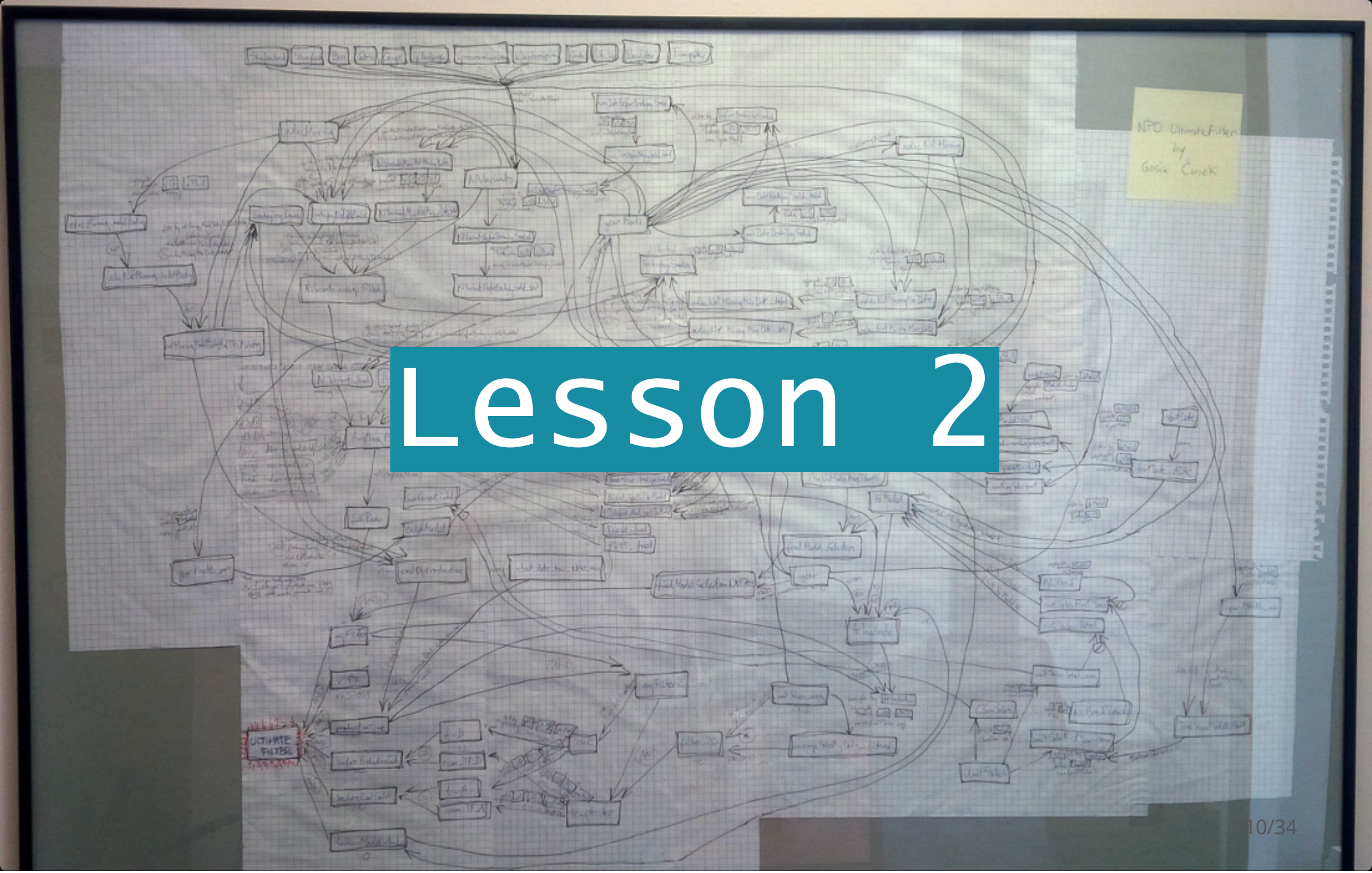

Lesson 1

What is a data scientist?

- **Data Science:**
- IT infrastructure + Databases + Data preparation + Data modelling + Visualization + Story telling
- **So you just need:**
- Mathematics, Statistics, Computer science, Machine Learning, Design of experiments, Bayesian analysis, Super Vector Machines, Linear Discrimination Analysis, Neural Networks, Naive Bayes Classifier, Regression Methods, Clustering Methods, Optimization, Gradient descents, Business Orientation, Hacker Mindset, Problem Solver, Python, R, SPSS, KNIME, C++, SQL, NoSQL, Parallel Computing, Linear Algebra, Calculus, MapReduce, Hadoop, Hive, Pig, Spark, Amazon Web Services, ggplot2, ggvis, D3.js, Story Telling, Presentation Skills, Reproducibility Tools, Markdown, Pandoc, Python notebooks, Shiny, Communication Skills, Version Control, Package Development, Coding Skills, H20, Big Data Orientation, Business Intelligence Tools, Weka, Network Analysis, Pandas, NumPy, SciPy, data.table, dplyr, MongoDB, MonetDB, and I will not continue because you see the point.


YEAH,
GOOD
LUCK
WITH
THAT
MISSING
"HORSEY THE UNICORN"
EYES: BLUE
HAIR: RAINBOW

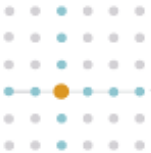

A data scientist is a team, not a person.



Lesson 2

Why do we love R?

- R vs SPSS
- R vs Excel
- R vs KNIME
- R vs Whatever point and click "analytics" software



Analyx®

Strategic Predictive Customer Insights

Why do we love R?

- R vs SPSS
- R vs Excel
- R vs KNIME
- R vs Whatever point and click "analytics" software

Reproducibility!

Analyx®

Strategic Predictive Customer Insights

What is reproducibility?

Your primary collaborator is yourself 6 months from now, and your past self doesn't answer emails

- Reproducibility is a process for sharing the methodology, describe the environment, and recreate results. (Andrie de Vries)
- For Academia: Verify results!
- For Business: Production code, reliability, reusability, collaboration, regulation.

So let's use R instead of Excel, and we are done?

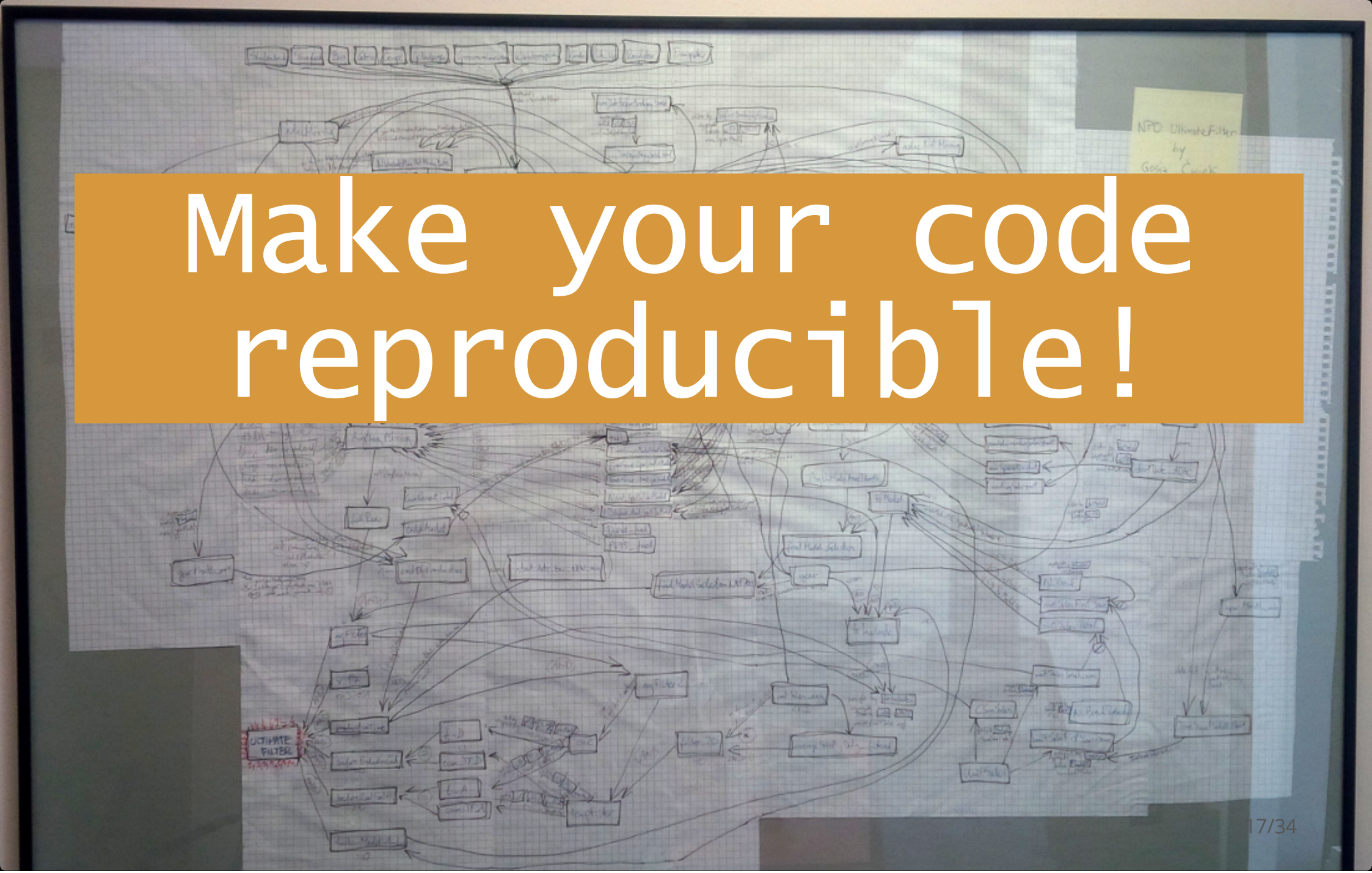
- When you finish your code everything looks like:

So let's use R instead of Excel, and we are done?

- But after some time you can easily end up with:

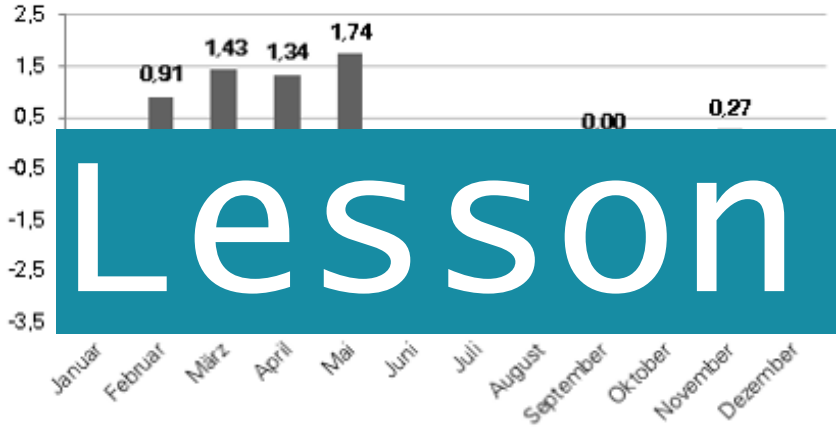
How to take care about reproducibility?

- Keep a consistent coding style
- Comment your code
- Write documentation
- Use version control systems
- Use Rstudio projects
- Organize your code into functions / packages


Make your code reproducible!

Model results

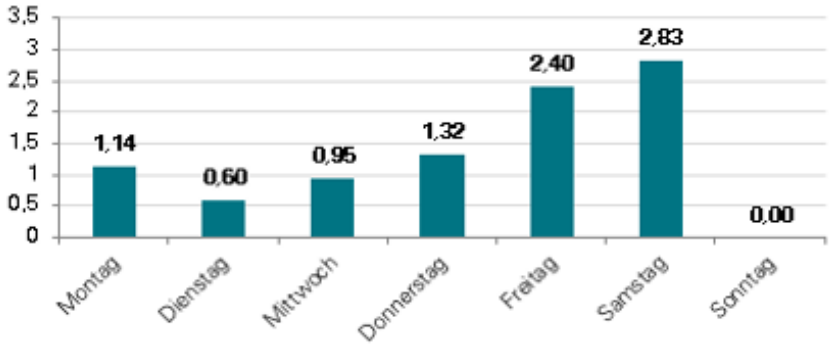
Seasonal effects



On average:

In **May** ceteris paribus consumers are willing to pay **1.74€ more** than in September

er ceteris paribus e willing to pay **2.35€** n in September



On **Saturday** ceteris paribus consumers are willing to pay **2.83€ more** than on Sunday

Lesson 3

We wanted to improve our outputs

First attempt: Google vis (demo)

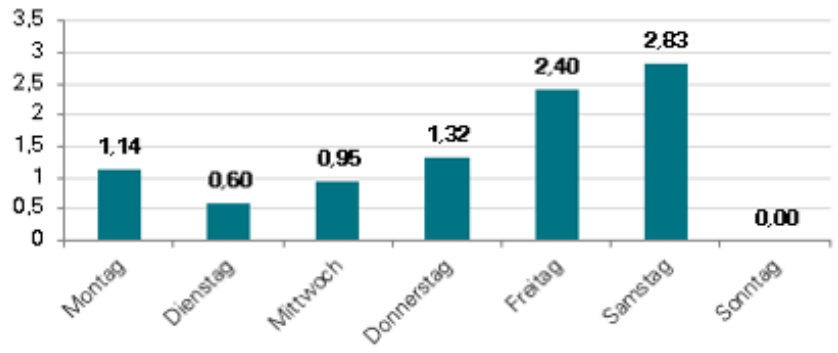
To give our customers control over the results

Second attempt: Shiny app (demo)

In an attractive way!

Third attempt: C# + R + web design + D3.js (demo)

Provide interactive solutions!



On **Saturday** ceteris paribus consumers are willing to pay **2.83€ more** than on Sunday

Analyx

brand.cision 35


Lesson 4

Once again, we had a problem

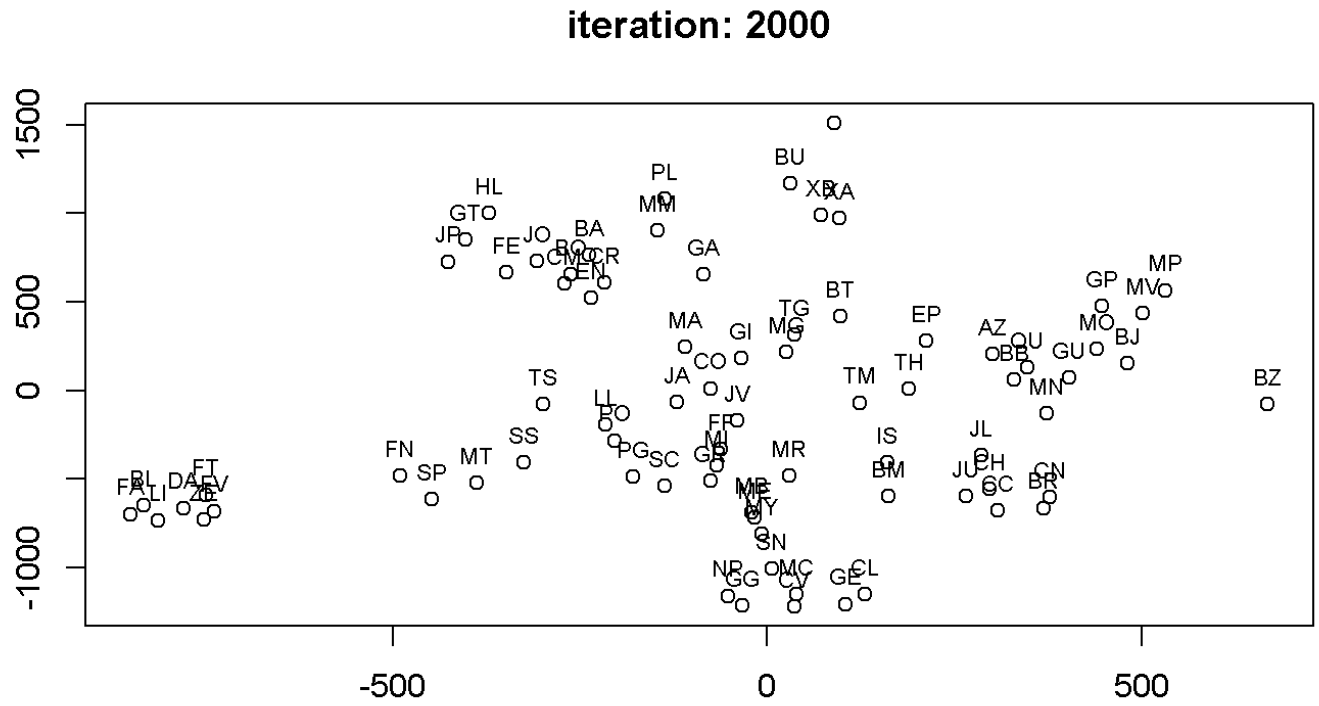
Social Network analysis example:

- Gephi + Excel
- Python?
- R + igraph
- R + igraph + own development

We took the best from others to build our own solution

```
#devtools::install_github("analyxcompany/ForceAtlas2")
library(ForceAtlas2)
library(igraph)
g <- nexus.get("miserables")
layout <- layout.forceatlas2(g, iterations=2000, plotstep=100)
```

We took the best from others to build our own solution

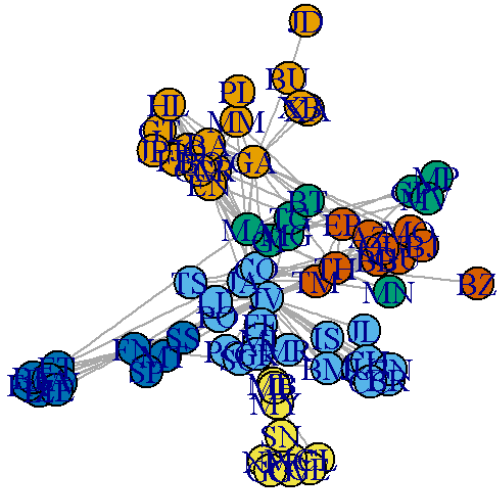


Strategic Predictive Customer Insights

We took the best from others to build our own solution

```
#devtools::install_github("analyxcompany/resolution")
library(resolution)
V(g)$color <- cluster_resolution(g)$membership
plot(g, layout=layout)
```

We took the best from others to build our own solution




Stand on the shoulders of giants

Lesson 5

Lesson 5

- We are intensive users of R and its packages
- dplyr, ggplot2, igraph, RandomForest, etc…
- We need to give back to the community!
- We (Analyx) regularly sponsor the Poznan R Users Community (PAZUR) meetings, and we sponsored the last Polish R Users community (also PAZUR)
- We open source code: The ForceAtlas2 and resolution packages

Contribute to the R community!

Summary

- Lesson 1: A data scientist is a team, not a person.
- Lesson 2: Make your work reproducible.
- Lesson 3: Provide interactive solutions.
- Lesson 4: Stand on the shoulders of giants.
- Lesson 5: Contribute to the R community.

Thank you, and keep contact!

- Web: www.analyx.com / www.adolfoalvarez.cl
- e-mail: adolfo.alvarez@analyx.com
- twitter: @analyxcompany / @adolfoalvarez