

Do French students cost too much?

Coding Week - Datathon
24.01.19

Francesco, Gloriana, India,
Max, Ricardo, Stephanie

We assessed the cost per student in two different ways

International comparison

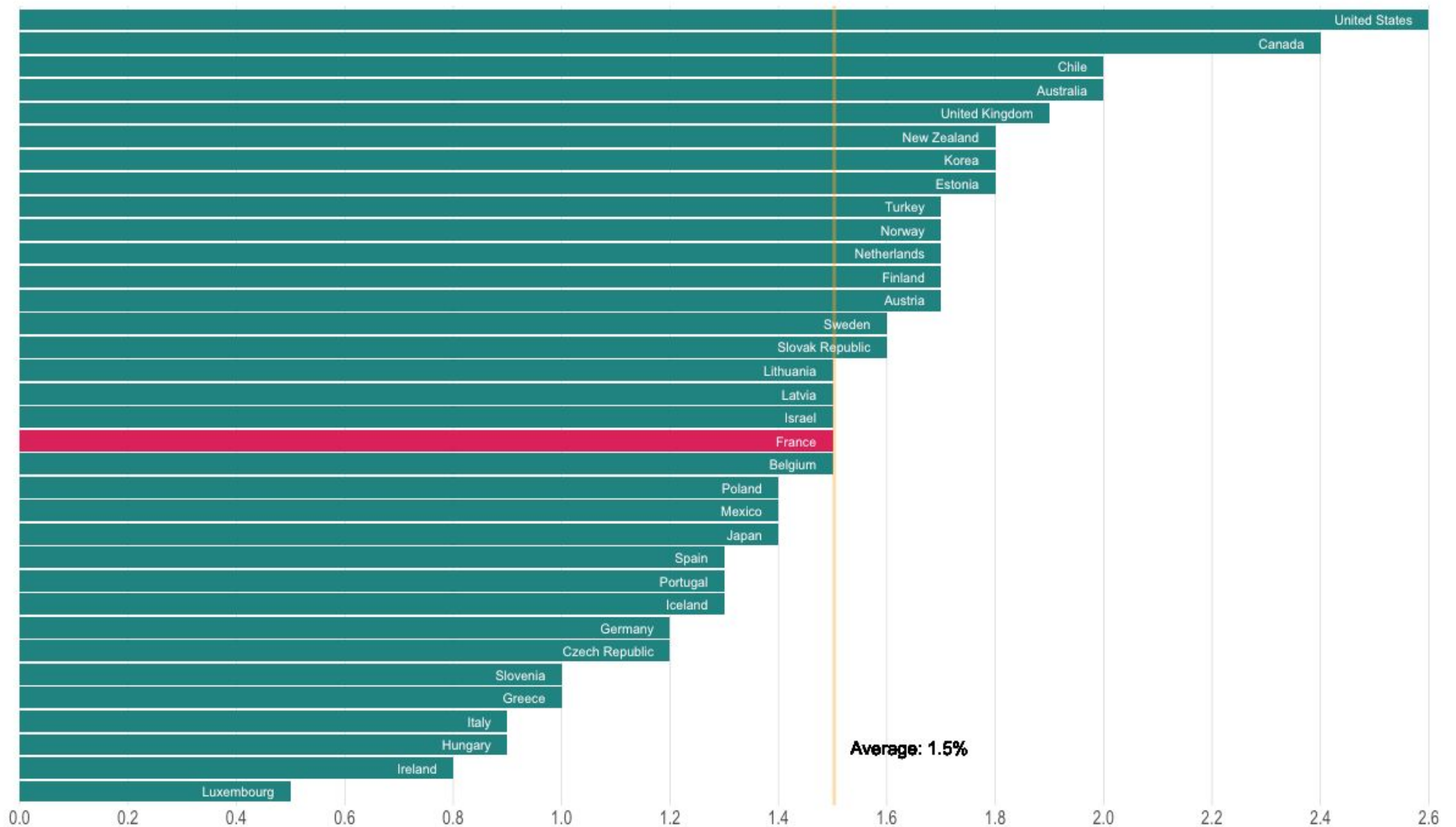
First, we used **OECD data** to check how much France spends in university student in comparison to other, similar countries.

Regional comparison

Second, we wanted to check whether spending per student within France varies across regions using data from the **Ministry of Education**.

Spending per student in France is average compared to other OECD countries

Education expenditure in OECD countries as % of GDP (2015)



~/Documents/school/sciences-po/datathon - master - RStudio

script.R* x frReg x tempmerged x regions x

Source on Save Run Source

read Next Prev All Replace Replace All

☐ In selection ☐ Match case ☐ Whole word ☐ Regex ☒ Wrap

```

65 # -----
66 # --- IMPORTING DATA SETS ---
67 #
68 # -----
69
70 # // object: eduper
71 # // expenditure as $ GDP (OECD)
72 edu0ecd <- read_csv("../data/OECD_StudentExpenditure-edited.csv")
73 edu0ecd <- na.omit(edu0ecd)
74 edu0ecd <- select(edu0ecd, 'country', 'allGdp')
75 # export data frame of % GDP as file
76 write.xlsx(edu0ecd, "../data/edu-oecd.xlsx")
77 view(edu0ecd)
78
79 # // temp object: eduregio (for edu)
80 # // Effectifs d'étudiants inscrits 2014-15
81 eduregio <- read_delim("../data/Enrollment.csv", delim = ";", locale = locale(encoding = "Latin1"))
82 view(eduregio)
83 studnum <- eduregio %>%
84   group_by(Région) %>%
85   summarise(studnum = sum(as.numeric(`Effectifs d'étudiants inscrits 2014-15`), na.rm = TRUE))
86
87 # // temp object: eduexp [for edu]
88 # // spending data
89 eduexp <- read_delim("../data/SpendingData.csv", delim = ";", locale = locale(encoding = "Latin1"))
90 view(eduexp)

```

Environment History Connections Git

Global Environment

- edu0ecd 34 obs. of 3 variables
- eduregio 312 obs. of 34 variables
- frMap List of 8
- frReg 13 obs. of 3 variables
- lab List of 8
- mapSpending 1009 obs. of 7 variables
- mergedMap Large mapview (3.1 Mb)
- nameSpend 13 obs. of 3 variables
- points 18 obs. of 2 variables
- regdata List of 3
- regions 18 obs. of 3 variables
- regionsAnd.. 13 obs. of 9 variables
- simpleMap 13 obs. of 3 variables
- tempfrMap Large mapview (3.1 Mb)
- tempmerged 169 obs. of 6 variables

Files Plots Packages Help Viewer

Zoom Export

Feature ID 158

code 24

2 nom Centre-Val de Loire

erStudent 278.05766547822

103216

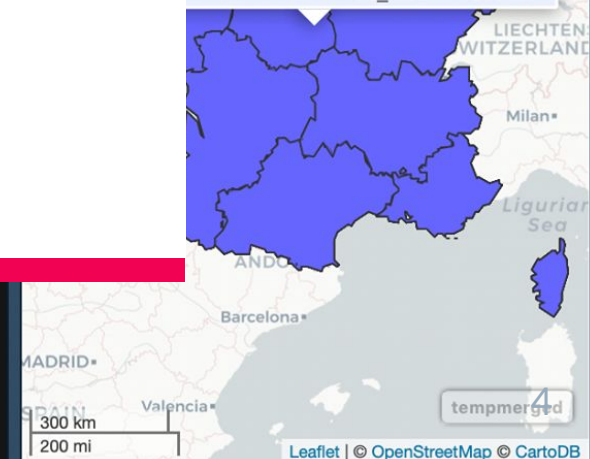
sfc_MULTIPOLYGON

Our hypothesis: Funding per student varies across regions

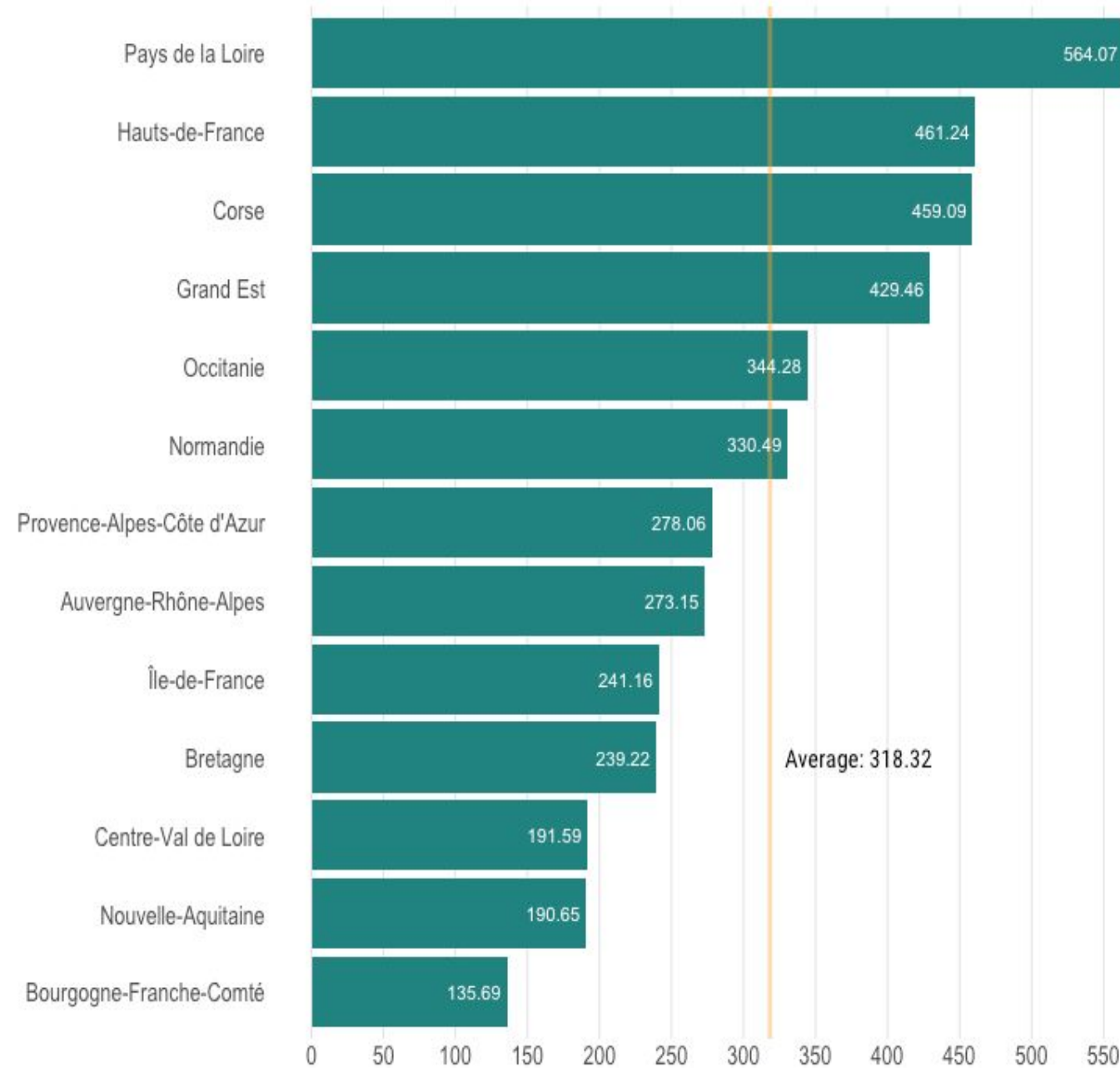
```

> tempmerged <- merge(frReg, nameSpend, all.x=TRUE,no.dups)
Error in fix.by(by.x, x) : object 'no.dups' not found
> tempmerged <- merge(frReg, nameSpend, all.x=TRUE,no.dups=TRUE)
> view(tempmerged)
> View(frReg)
> |

```

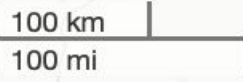
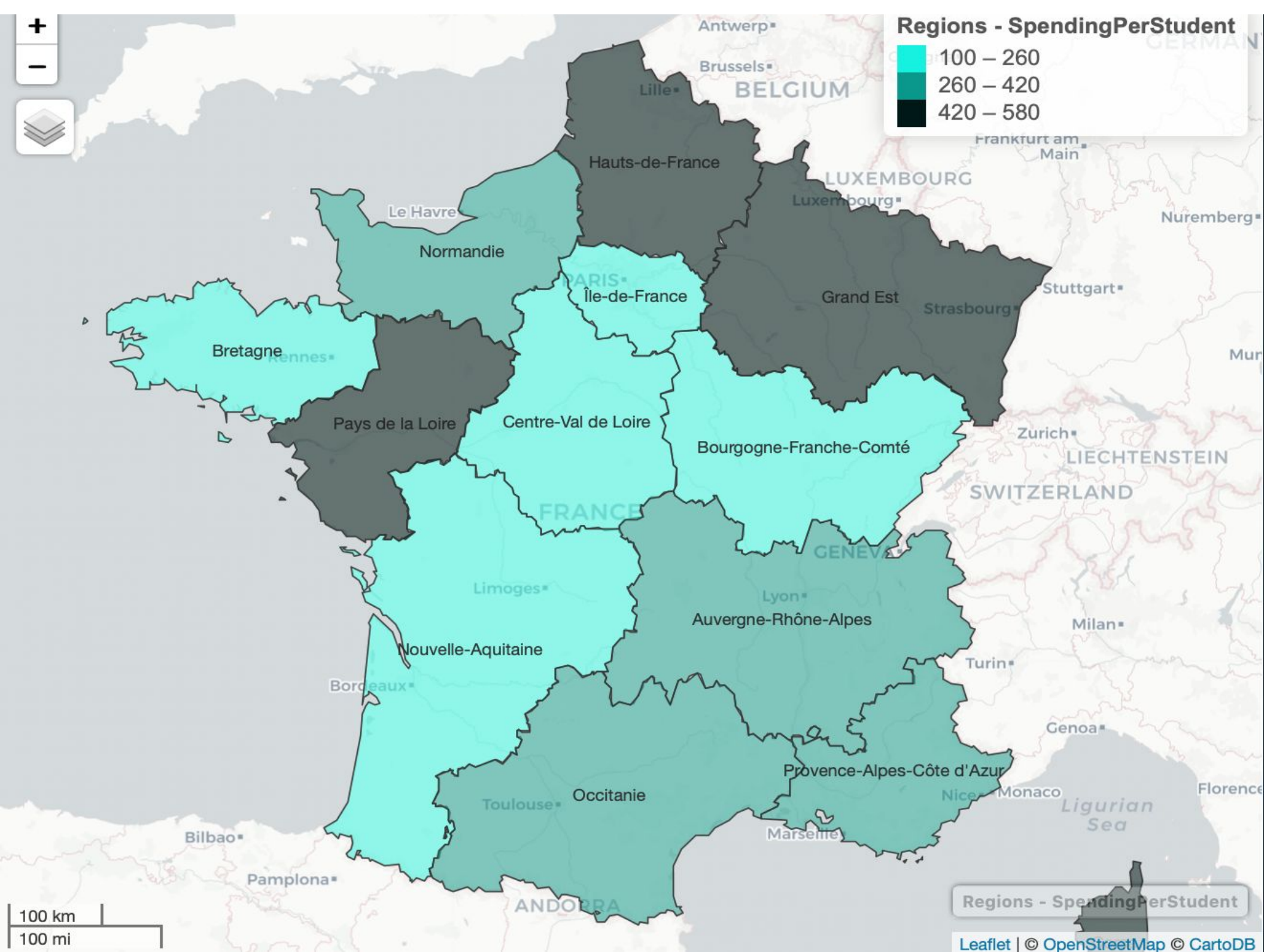


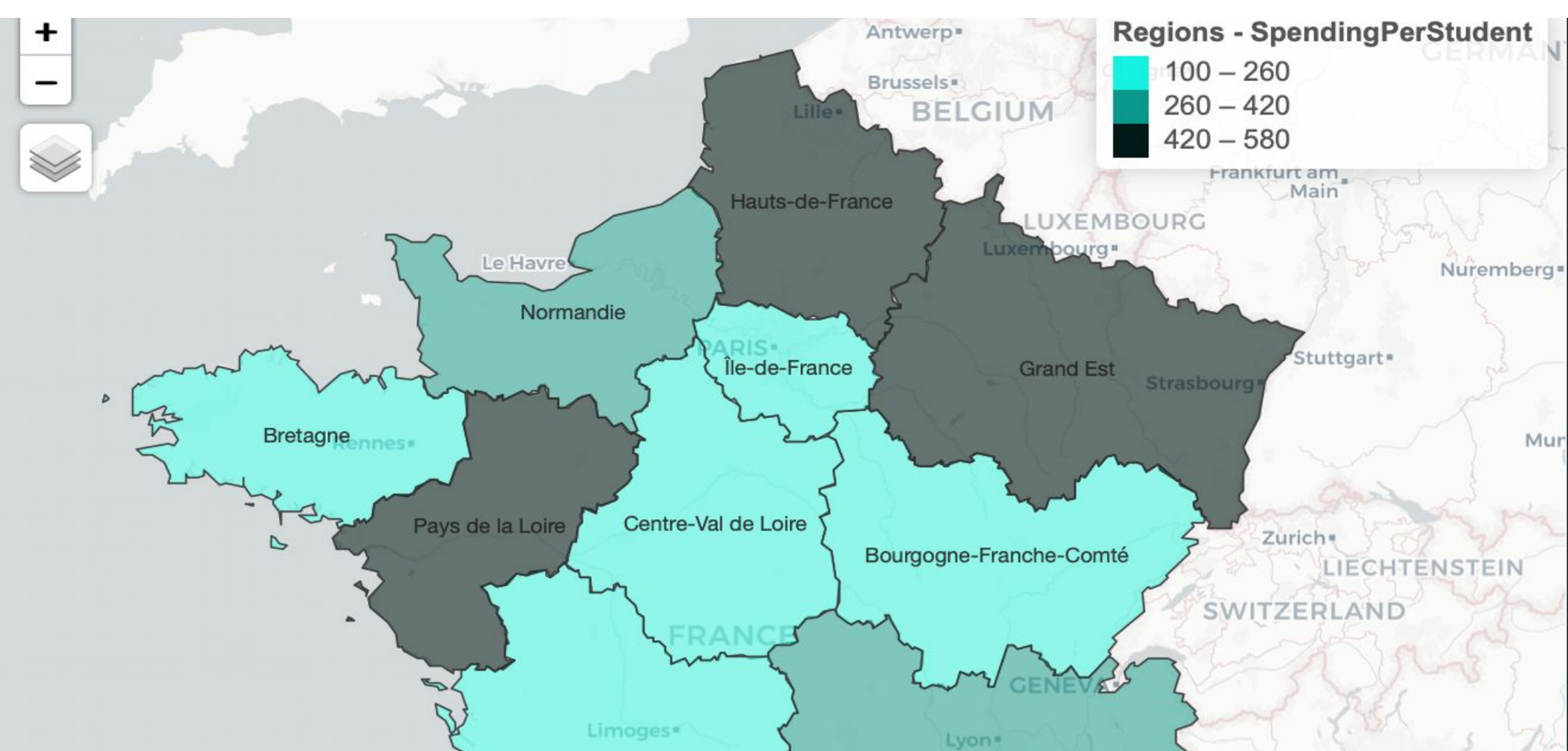
Spending per student in each region in France



Source: Ministère de l'Éducation nationale et de la Jeunesse, computation by Sciences Po students.

There is a lot of variation in spending per student - but it's hard to find discernable patterns





BUT: Take these results with a few grains of salt.



THE DATA IS DARK AND FULL OF ERRORS



*Lord of Light protect us,
for the night is dark and full of terrors!*

We were faced with various challenges with our data



Lots of missing data in the datasets we used



Very little and bad data available on HE spending



Some data was only available through a terrible UI

Effectifs d'étudiants inscrits 2012-13	Effectifs c
470	455
736	717
12820	12705
nd	nd
25020	26120
2715	2953
1333	1280
3530	3617
28397	28766
17174	17383
528	570
Comptabilisés dans les effectifs de leur établis	Comptabilisé
7344	7331
651	663
1130	1304
933	981
424	429

We were faced with various challenges with our data



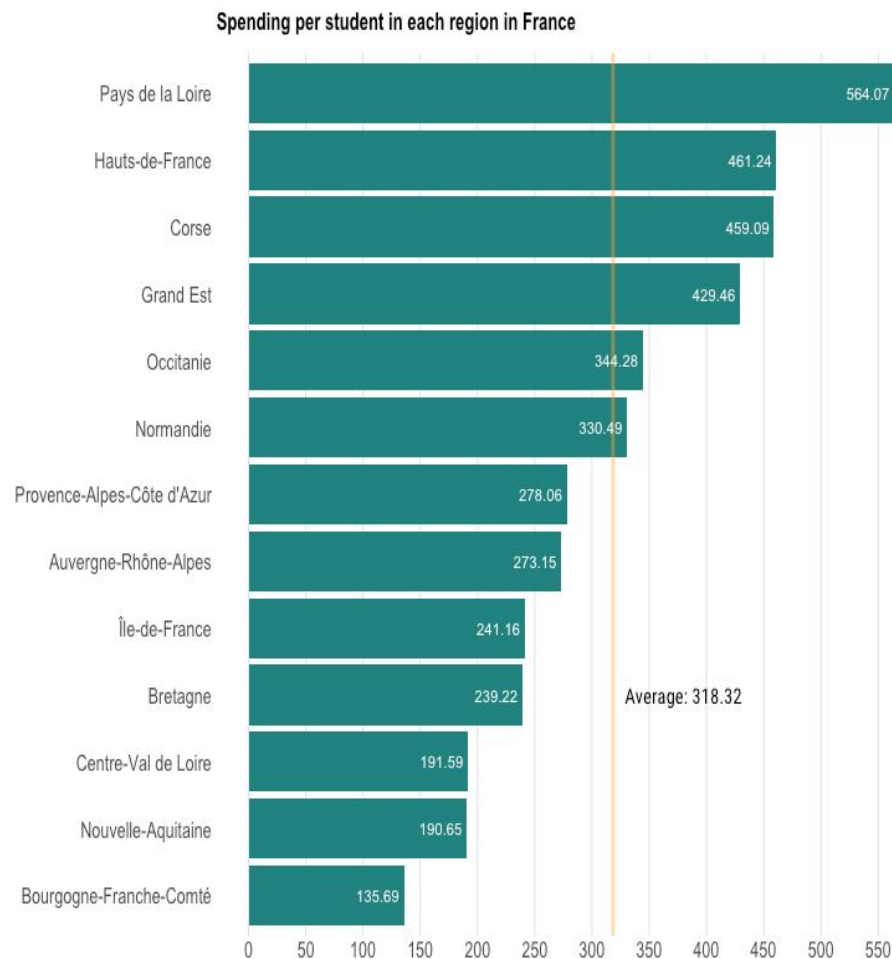
Lots of missing data in the datasets we used



Very little and bad data available on HE spending



Some data was only available through a terrible UI



Source: Ministère de l'Éducation nationale et de la Jeunesse, computation by Sciences Po students.

We were faced with various challenges with our data



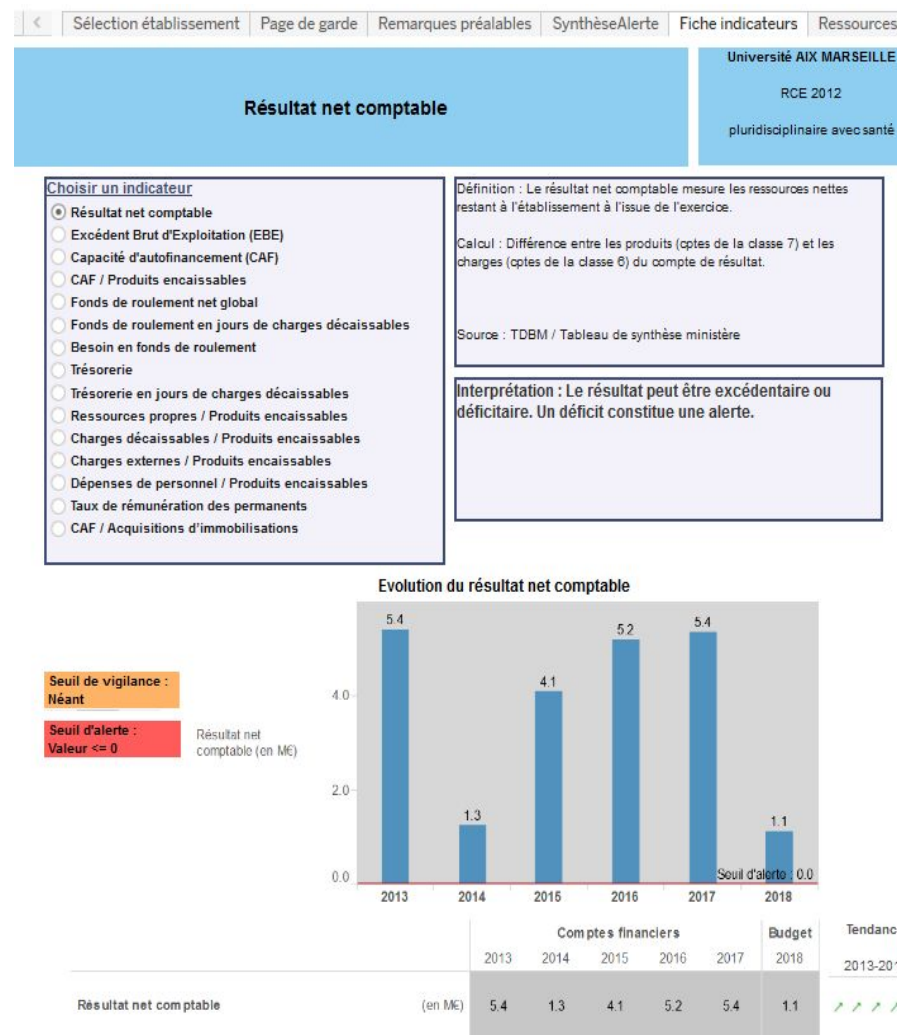
Lots of missing data in the datasets we used



Very little and bad data available on HE spending

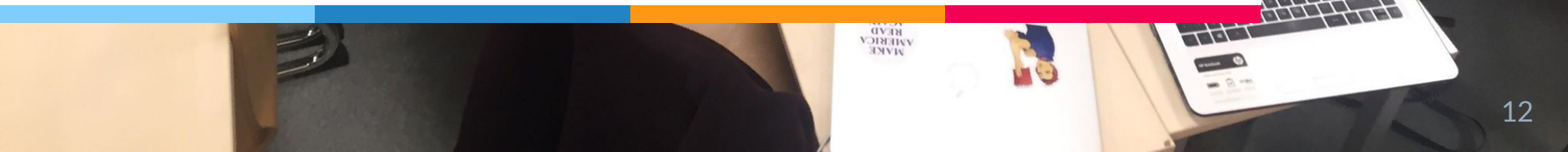


Some data was only available through a terrible UI





This is why by yesterday evening we looked like this



France may be at the forefront of the open data movement - but there's still a lot of work left

So how can we address the issues we encountered?

Create an incentive system for public org's

To benchmark ministries and public agencies, we suggest an open data score based on **usability, relevance and completeness**.

Make quality control a high-level task

To enforce standards and **ensure data quality**, there needs to be a high-level unit in charge of controlling and curating datasets.

Take user feedback seriously

To effectively use the feedback data users provide, a **rating system** can be used to flag bad datasets and trigger consequent actions.

Thanks!

Any questions?

Coding Week - Datathon

24.01.19

Francesco, Gloriana, India,
Max, Stephanie, Ricardo