Using Big Data to Study Geographical Variation in Antibiotic Prescription

SPEAKER:

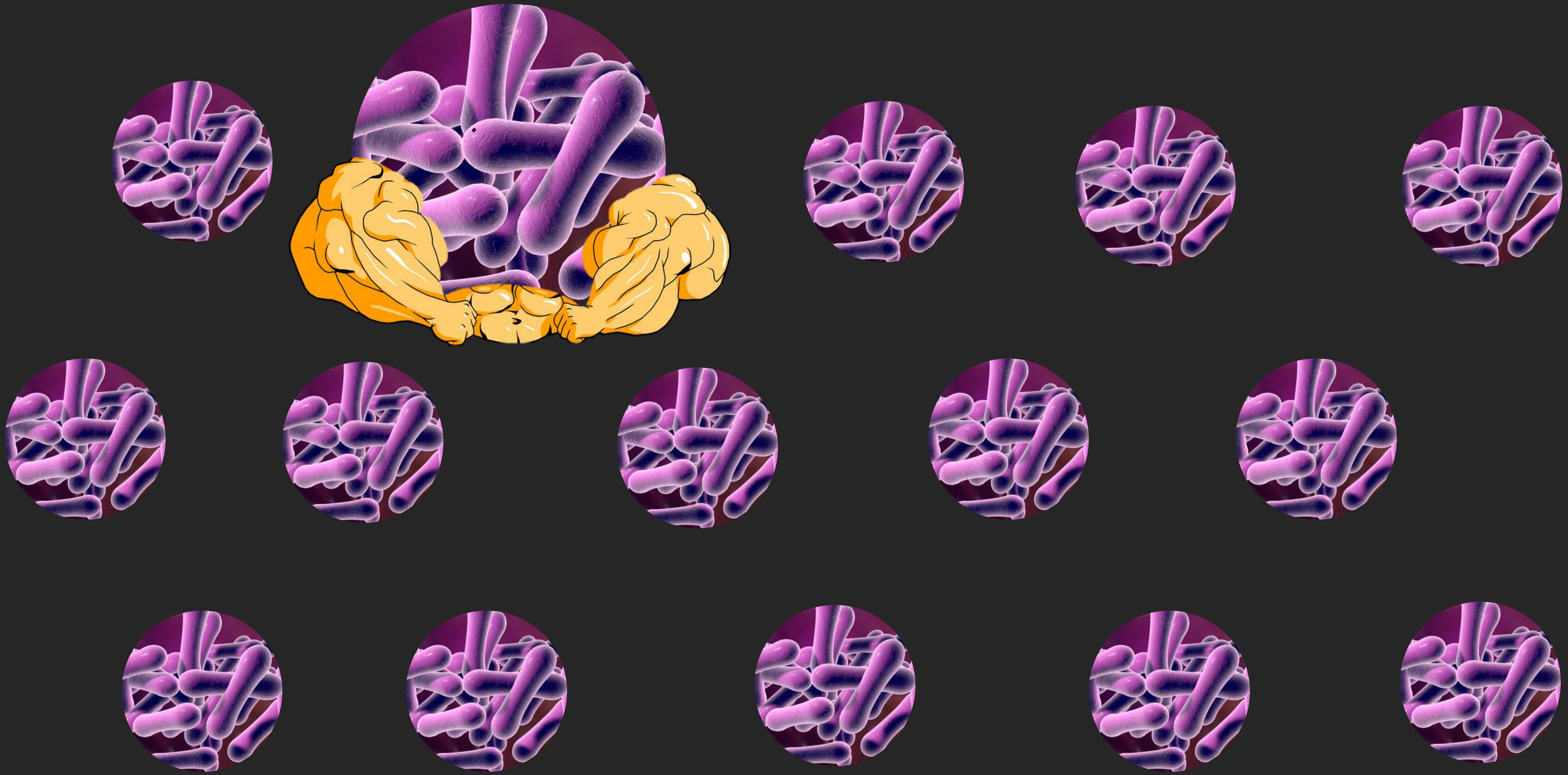
ARMANDO GONÇALVES

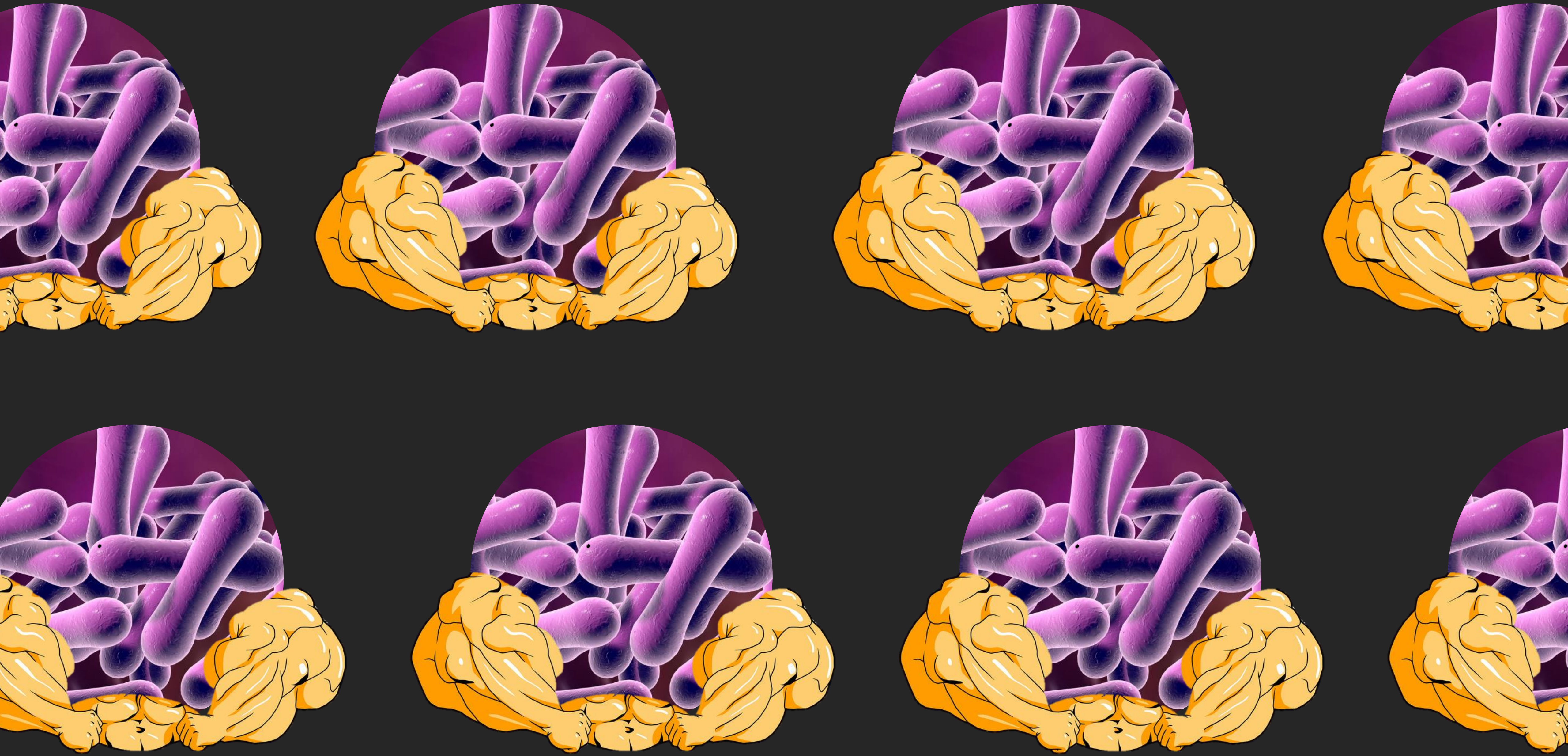
MENTORS:

IRMA VARELA, SARA MESQUITA

I.

Bacterial resistance





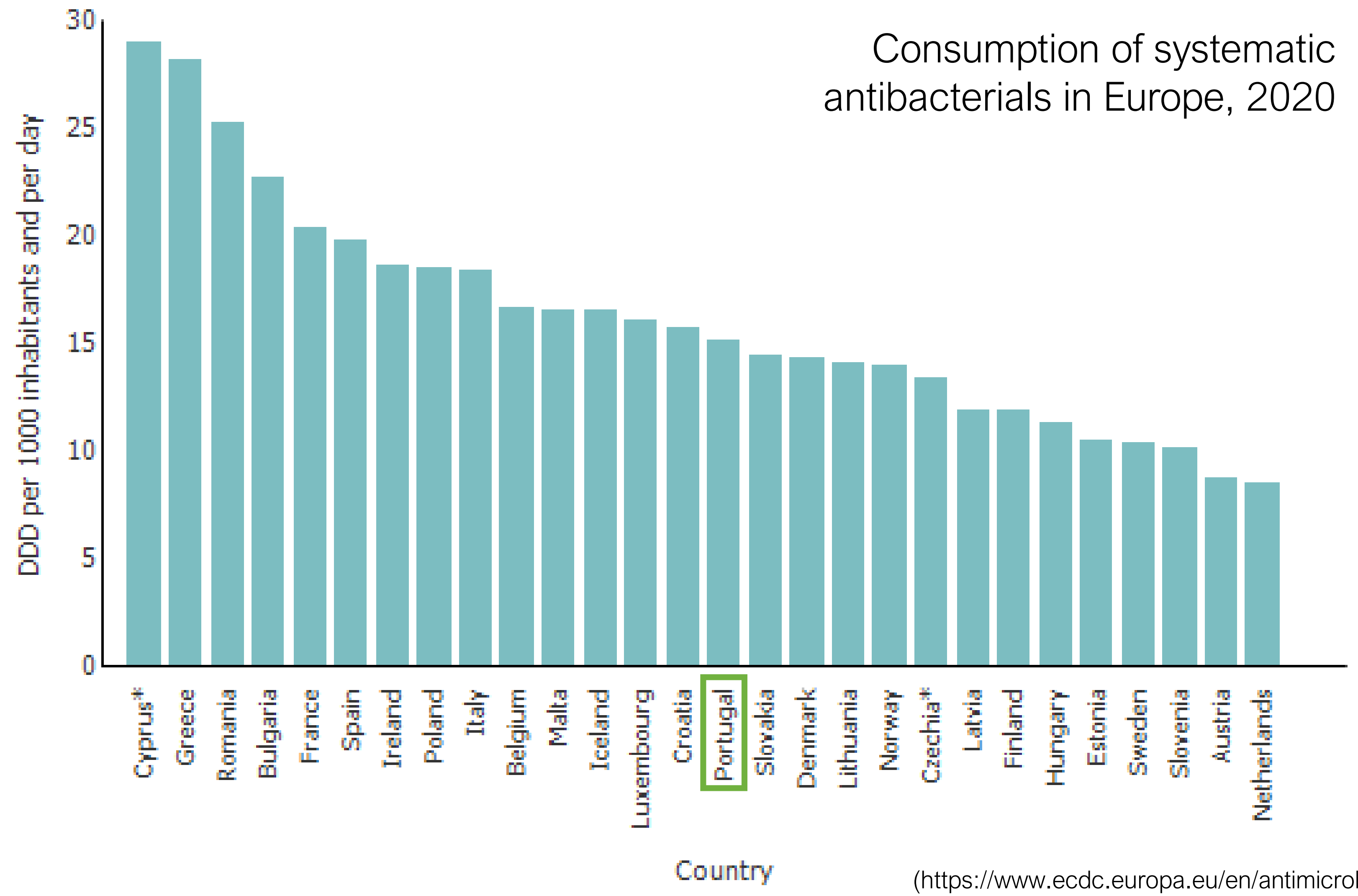

RIP
Tracy McConnell
Cause of death: otitis
(1984-2024)

Are there any **geographic** differences in antibiotic prescriptions in Portugal?

If there are indeed geographic differences, can we give it a **meaning**?

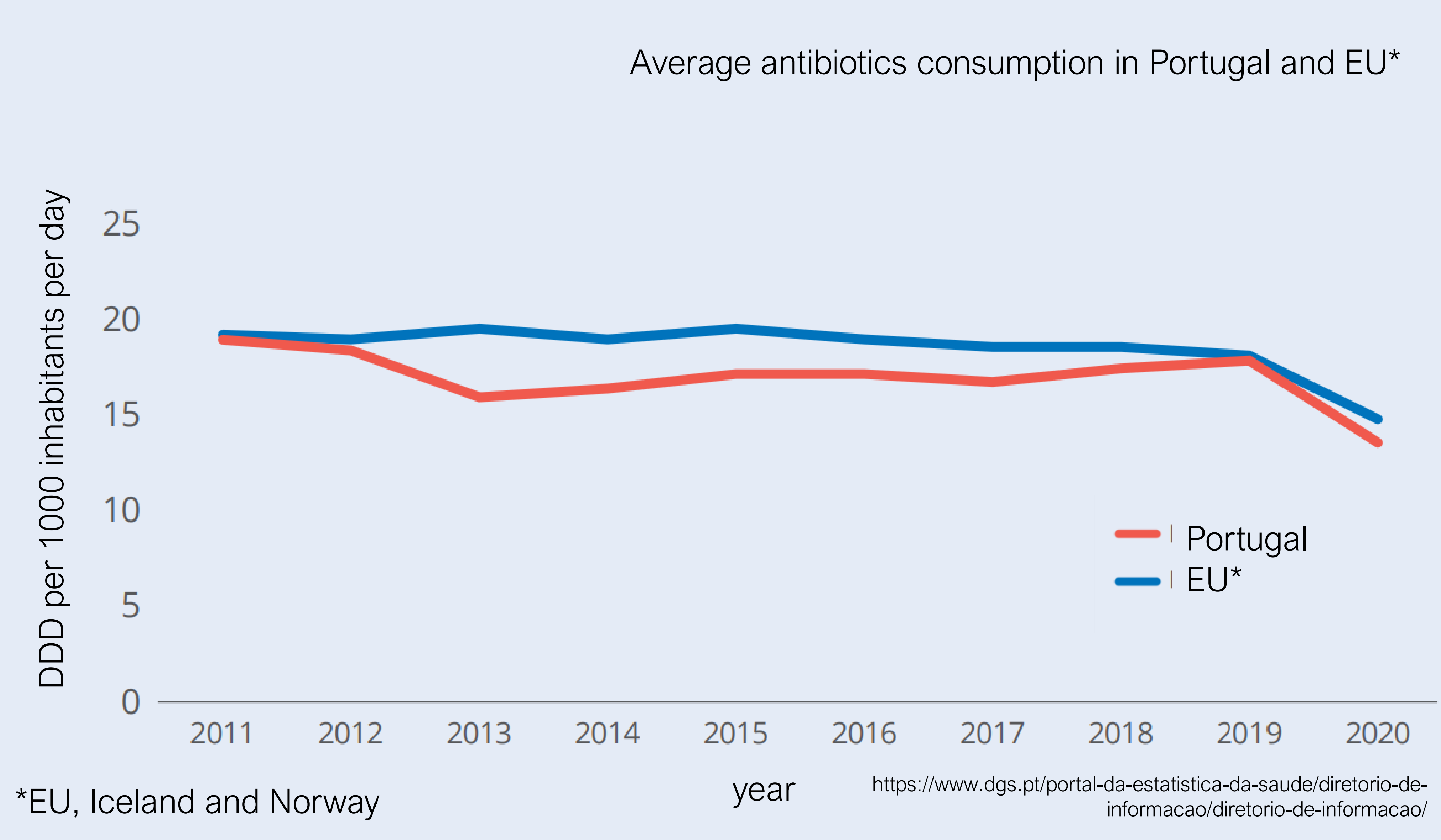
II.

Portugal’s case



* Country provided only total care data.

(https://www.ecdc.europa.eu/en/antimicrobial-consumption/database/rates-country)


Average antibiotics consumption in Portugal and EU*
DDD per 1000 inhabitants per day
25
20
15
10
5
0
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
year
Portugal
EU*
*EU, Iceland and Norway
https://www.dgs.pt/portal-da-estatistica-da-saude/diretorio-de-informacao/diretorio-de-informacao/

III.

The database

What does the database contain?

Medical prescriptions!

Lots and lots of prescriptions!...

https://www.spms.min-saude.pt

https://pem.min-saude.pt

All Portugal’s medical prescriptions from 2017 to 2019 in one place…

Locations

Location_id
Region
District
Municipality
Establishment_type
is_NHS

Prescriptions

Table_id
Prescription_id
Presc_date
Presc_time
Location_id
Patient_id
Patient_gender
Patient_age

Active substance
Dosage form
Dosage
Package type
Posology
Quantity
Prescriber_id
Speciality

Prescribers/MDs

Prescriber_id
Precriber_YOB
Prescriber_gender

2017 only…

	location_id	region	district	municipality	establishment_type	is_sns	service_type
	A95B172DF70598CF4328BC19232424A7V01	Lisboa Vale Tejo	Lisboa	Vila Franca de Xira	ACES-UCSP-Unid.Cuid.Saúde Personalizados	1	None
	5783455EABBB031808C10F9C88B7CA28V01	Lisboa Vale Tejo	Lisboa	None	ACES-UCSP-Unid.Cuid.Saúde Personalizados	1	None
	83C3E256FA9F1A8BCB6DEBA4476B5624V01	Lisboa Vale Tejo	Lisboa	Vila Franca de Xira	ACES-UCSP-Unid.Cuid.Saúde Personalizados	1	None
	01D318854A0A3094F791949124				ACES-USP-Unid.Saude Publica	1	None
					ACES-UCSP-		

Microsoft SQL Server Management Studio

File is too large to open.

OK

	patient_nid	patient_gender	p	substance	presc_date	presc_time
0	698959	F		urosemida	2019-12-31	10:46:44
1	698959	F	83	Brometo de aclidínio + Formoterol	2019-12-31	10:46:44
2	698959	F	83	Amoxicilina + Ácido clavulânico	2019-12-31	10:47:42
3	2627842	M	71	Pentoxifilina	2019-12-31	09:47:09
4	2627842	M	71	Clonazepam	2019-12-31	09:47:09

IV.

The work done


python ™

```python
PGHOST = 'localhost'
PGDB = 'pem' #
PGUSER = 'armando'
password = '[redacted]'
PGPASSWORD = password
print('Postgres password: ' + password)
```

Python

```python
dbconfig = {
    'port':5432,
    'host':PGHOST,
    'database': PGDB,
    'user': PGUSER,
    'password': PGPASSWORD,
}
```

Python

```python
engine = db.create_engine('postgresql://',connect_args = dbconfig)
```

Python

```python
connection = engine.connect()
Session = sessionmaker(bind=engine)
session = Session()
```

Python

```python
Table('locati ... load=True, autoload_with=engine)
ns location ...
```

Extraction → Storing data

```python
Dados_2017 = pd.read_csv('df_merged_2017.csv')
                         #parse_dates=['date'],
                         #index_col='date')
Dados_2017["metrics"] = ((((Dados_2017['#Ab_prescriptions']*100)/Dados_2017['#patient_AB']
#general standardised rate (GSR)
#Dados_2017["metrics1"] = ((((Dados_2017['#patient_AB']*100)/Dados_2017['TOTpatients'])*Da

dummy = Dados_2017.loc[:, ((Dados_2017.columns != 'municipality') & (Dados_2017.columns !=                                riptions


coco1 = dummy[['patient_gender', 'AgeGroup','TOTpatients']].groupby(['patient_gender', 'Ag

coco1.rename(columns = {'TOTpatients':'#TpatientsC'}, inplace = True)#the count was stored in patient_nid, let's change its name then
#coco1
```

[38]

Python

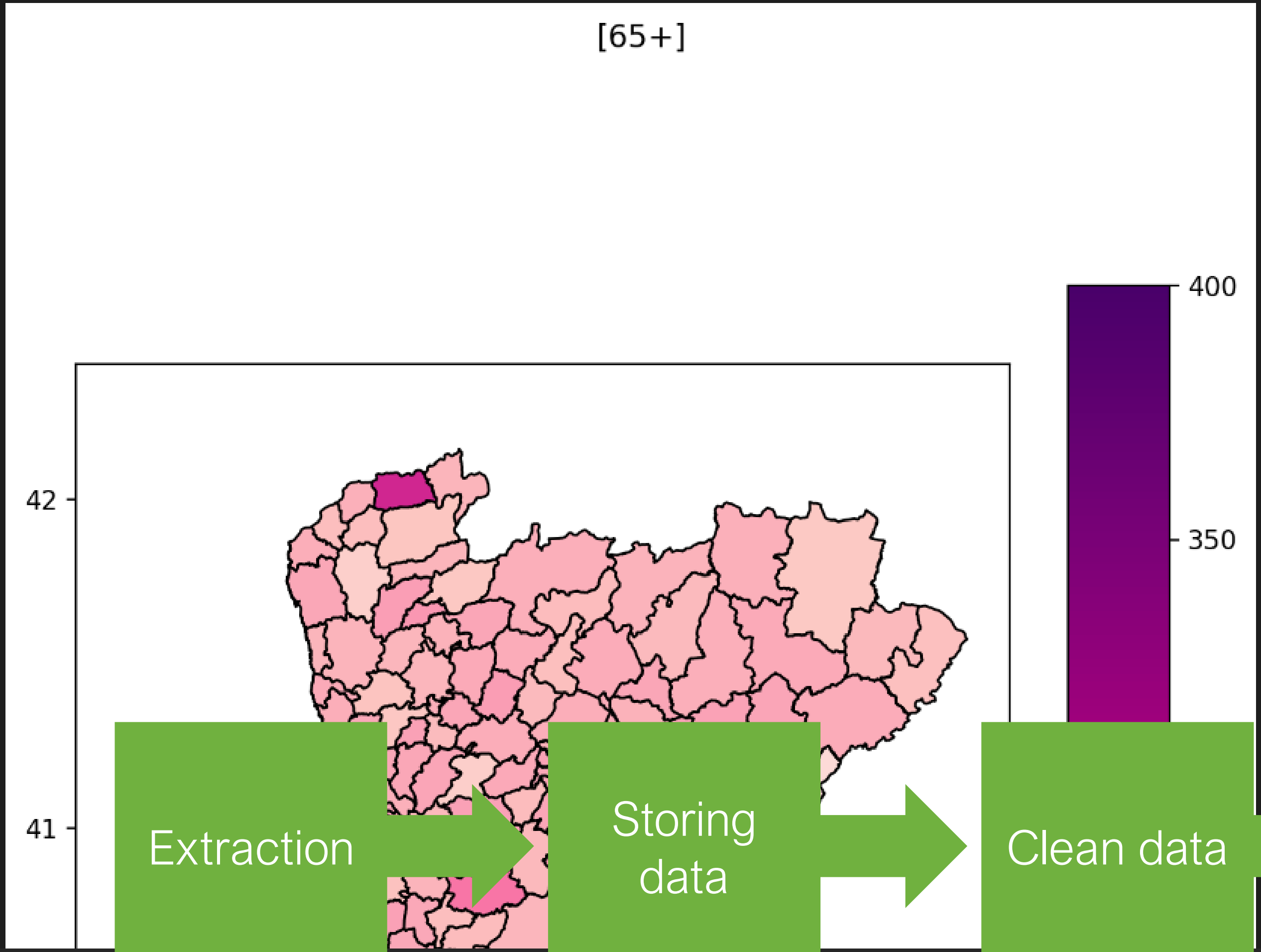
	patient_gender	AgeGroup	municipality	#Ab_prescriptions	#non_Ab_prescriptions	TOTpatients	#patient_non_AB	#patient_AB	metrics
119	F	[0-5]	Lisboa	47147.0	128709.0	53548	40633	27103.0	131.999514
424	F	[6-12]	Lisboa	23282.0	83303.0	41447	31818	15811.0	113.042232
729	F	[13-17]	Lisboa	15227.0	72067.0	30881	24683	10458.0	116.378377
1034	F	[18-24]	Lisboa	34603.0	135880.0	58084	45079	23535.0	114.108319
1339	F	[25-34]	Lisboa	62098.0	310009.0	107320	87546	42024.0	120.541299
1644	F	[35-44]	Lisboa	80518.0	487812.0	146111	123304	53168.0	127.801781
1949	F	[45-54]	Lisboa	66352.0	523443.0	136820	119218	43122.0	134.074865
2254		[55-64]		4433.0		136953		40493.0	142.424122
2559		[65+]		0160.0		275 18		86068 0	157.325232
2864									36.468540
3169		[6-12]		2794.0		47 75		16007	114.422130
3474		[13-17]		3243.0		29001		9218.0	114.888062

Extraction → Storing data → Clean data → Do the calculations →

```python
del ax
f,ax = plt.subplots(1,figsize=(8,14))
ax.set_axis_on()
f.suptitle('[0-5]')
ax = Data05F.plot(ax=ax,facecolor='blue', alpha=1,vmin = 0, vmax=400, edgecolor='black'
#plotty.plot(column='GSR', legend=True)
ax.figure.savefig('05F.png')
```

[93]

GeoPandas



Extraction

Storing data

Clean data

Do the calculations

Plot!!

Are there any **geographic** differences in antibiotic prescriptions?

If there are indeed geographic differences, can we give it a **meaning**?

V.

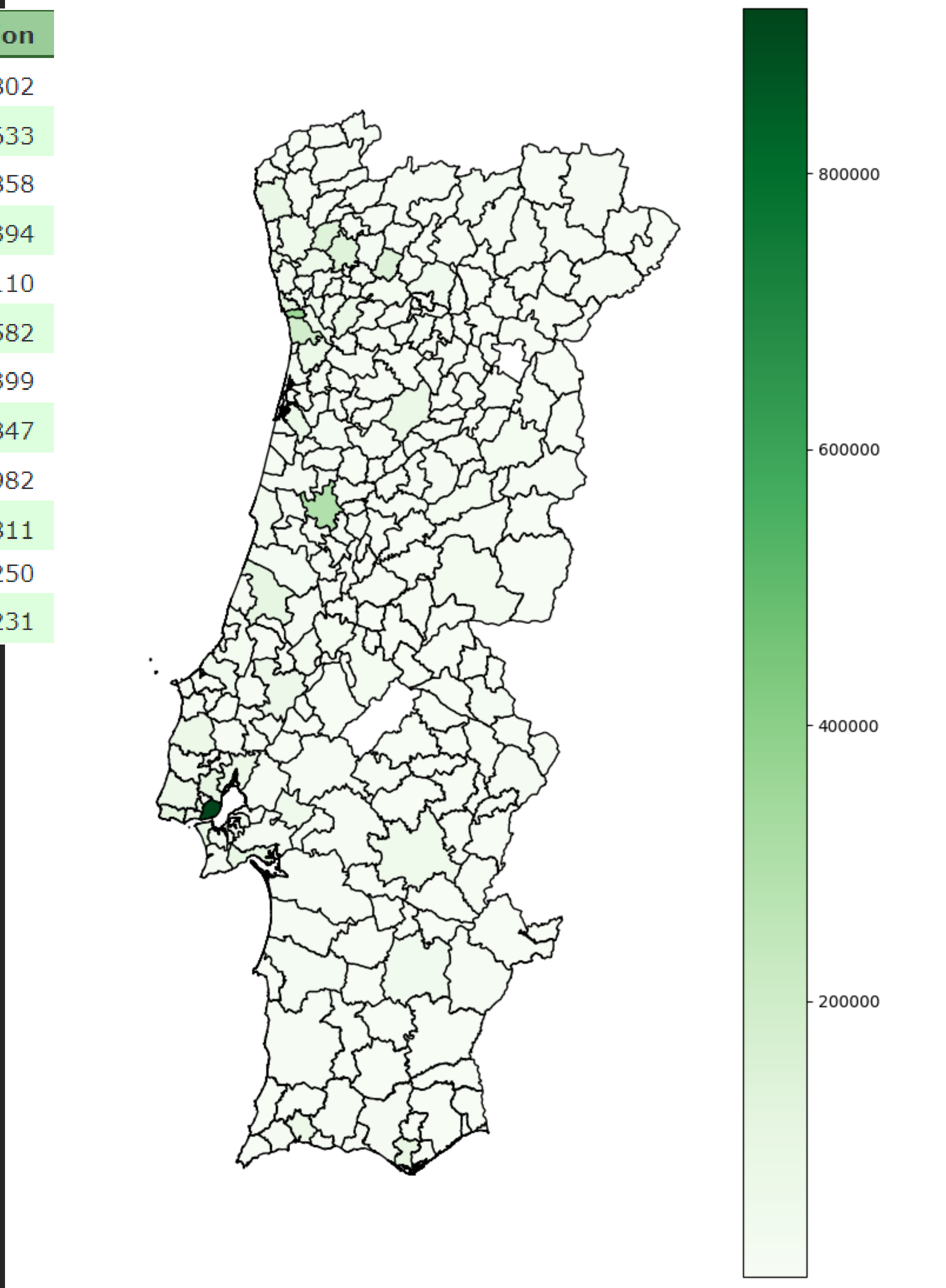
Methods and results

Babysteps…

Metrics #1:

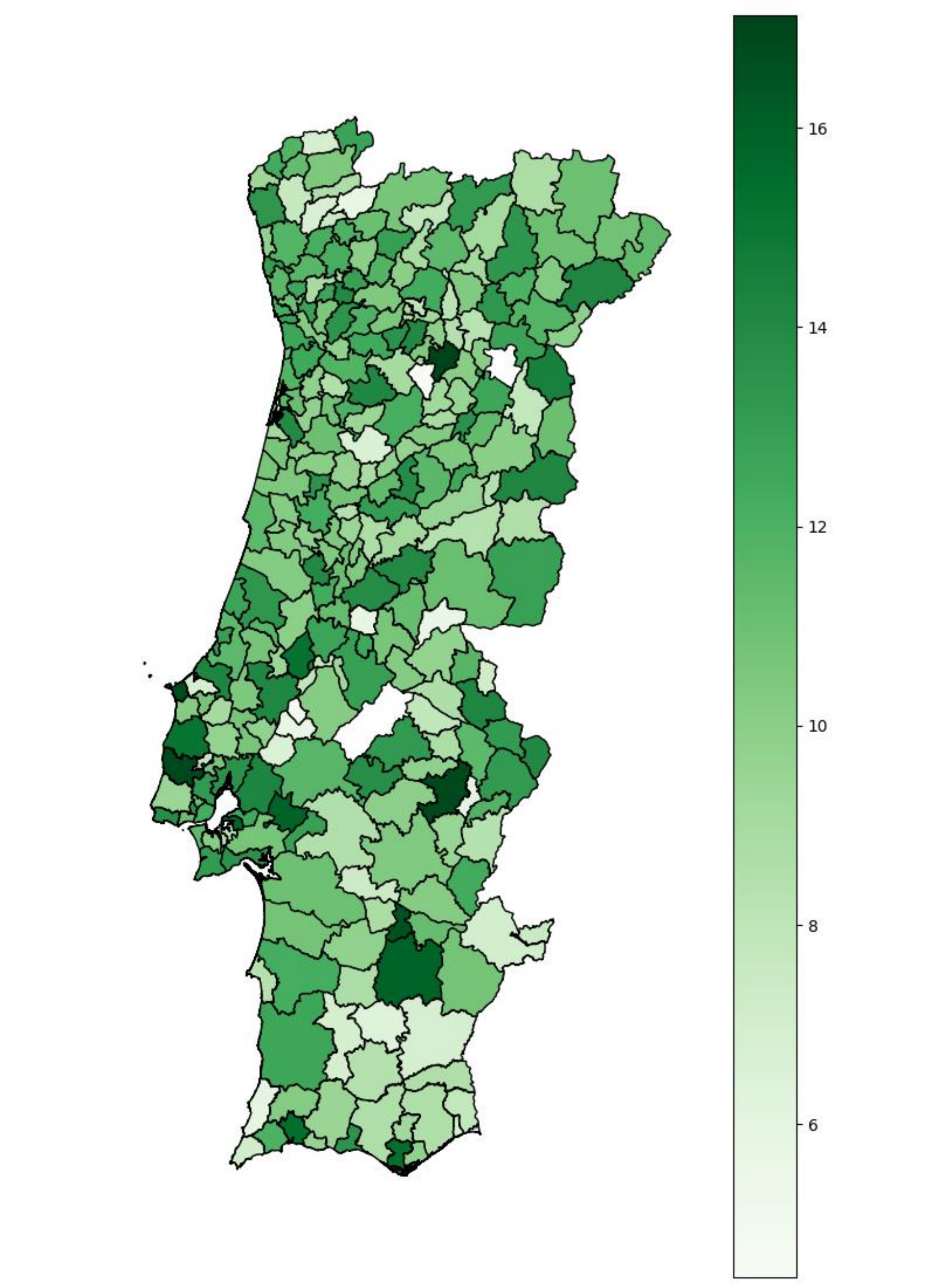
number of antibiotic prescriptions in the municipality

	Name	Population
1	Lisbon, Lisbon	517,802
2	Porto, Porto	249,633
3	Amadora, Lisbon	178,858
4	Braga, Braga	121,394
5	Setúbal, District of Setúbal	117,110
6	Coimbra, Coimbra	106,582
7	Queluz, Lisbon	103,399
8	Funchal, Madeira	100,847
9	Cacém, Lisbon	93,982
10	Vila Nova de Gaia, Porto	70,811
11	Algueirão, Lisbon	66,250
12	Loures, Lisbon	66,231



Metrics #2:

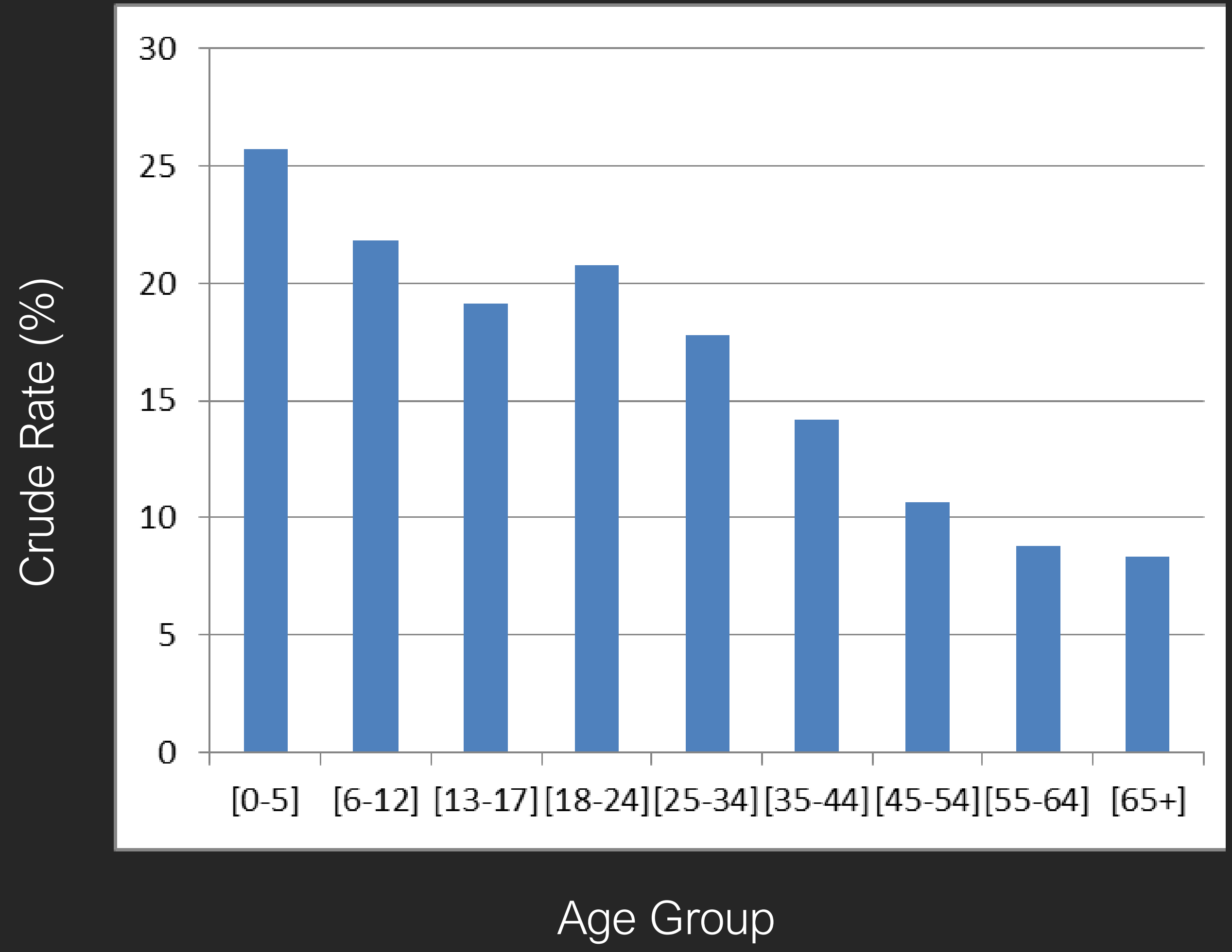
$$\frac{\text{number of antibiotic prescriptions in the municipality} * 100}{\text{TOTAL number of prescriptions in the municipality}}$$

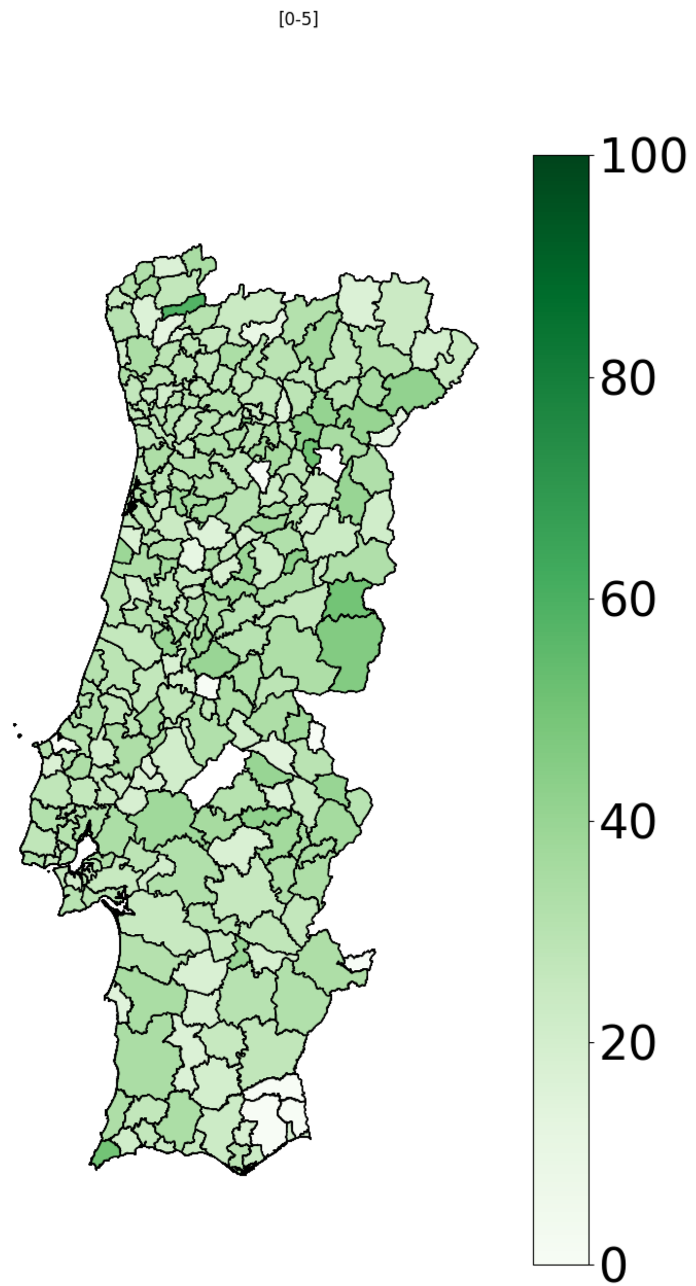

16
14
12
10
8
6

Metrics #2.1:

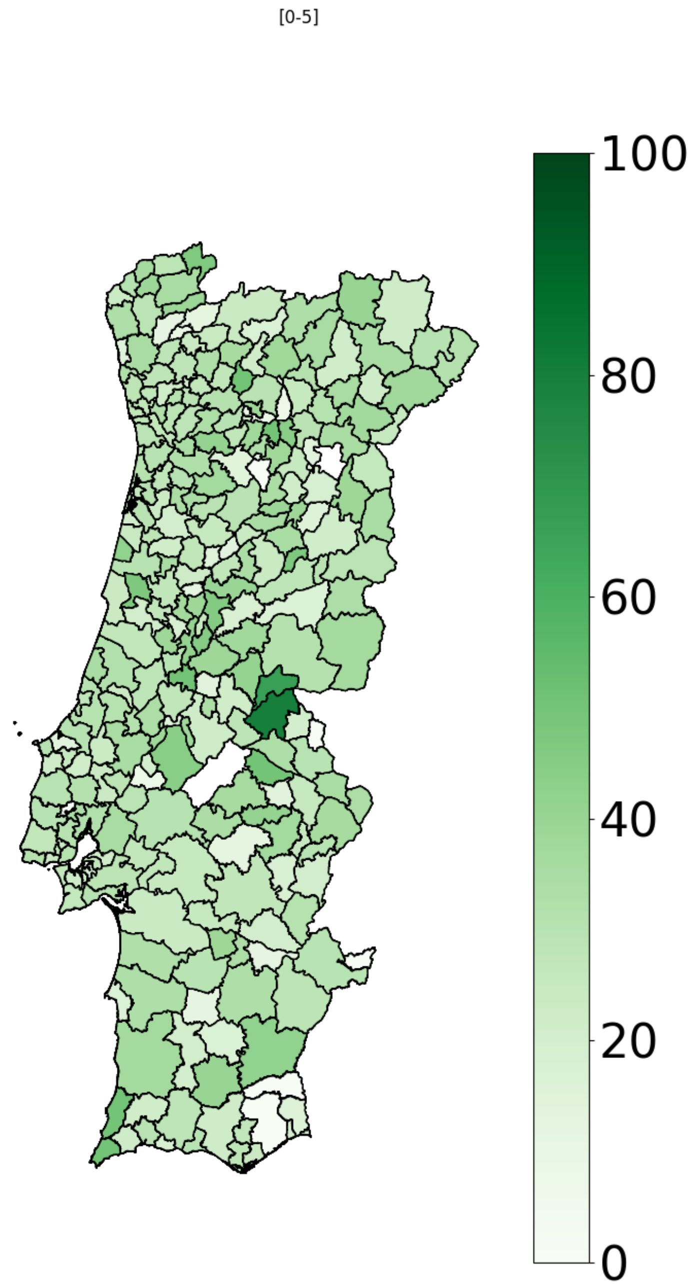
$$\frac{\text{number of antibiotic prescriptions in the municipality for specific age and sex} \; * \; 100}{\text{TOTAL number of prescriptions in the municipality for specific age and sex}}$$

Age Group vs Crude Rate

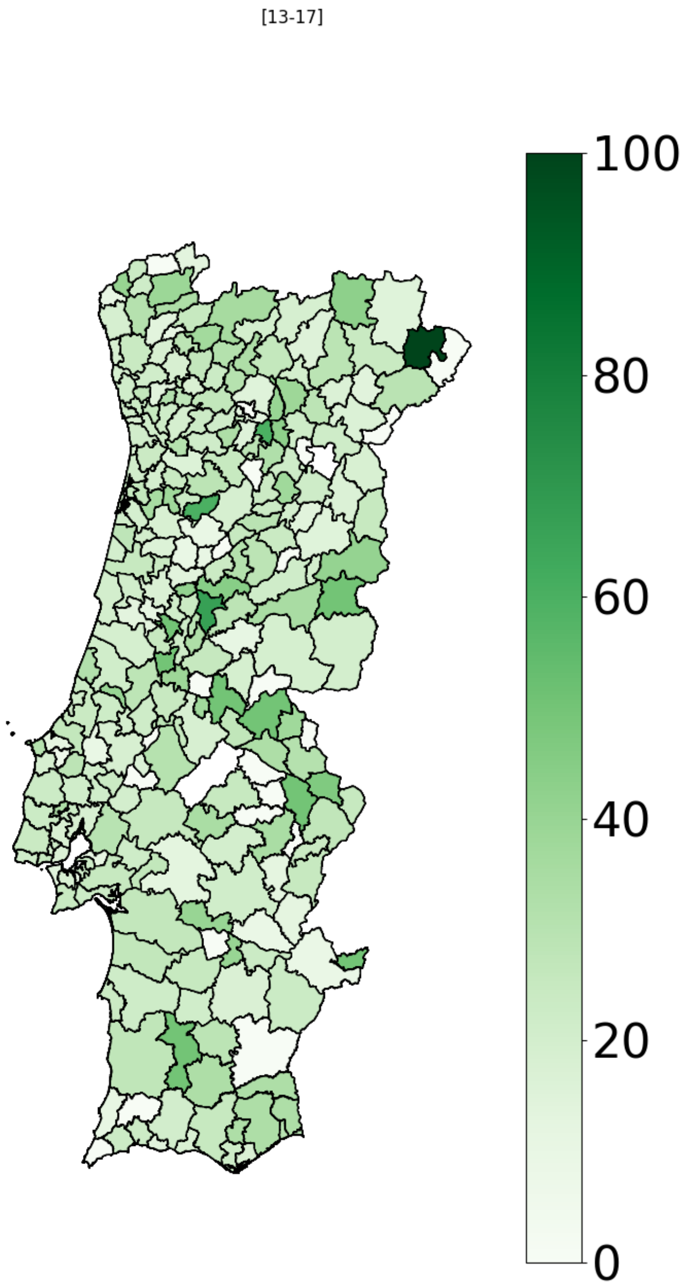


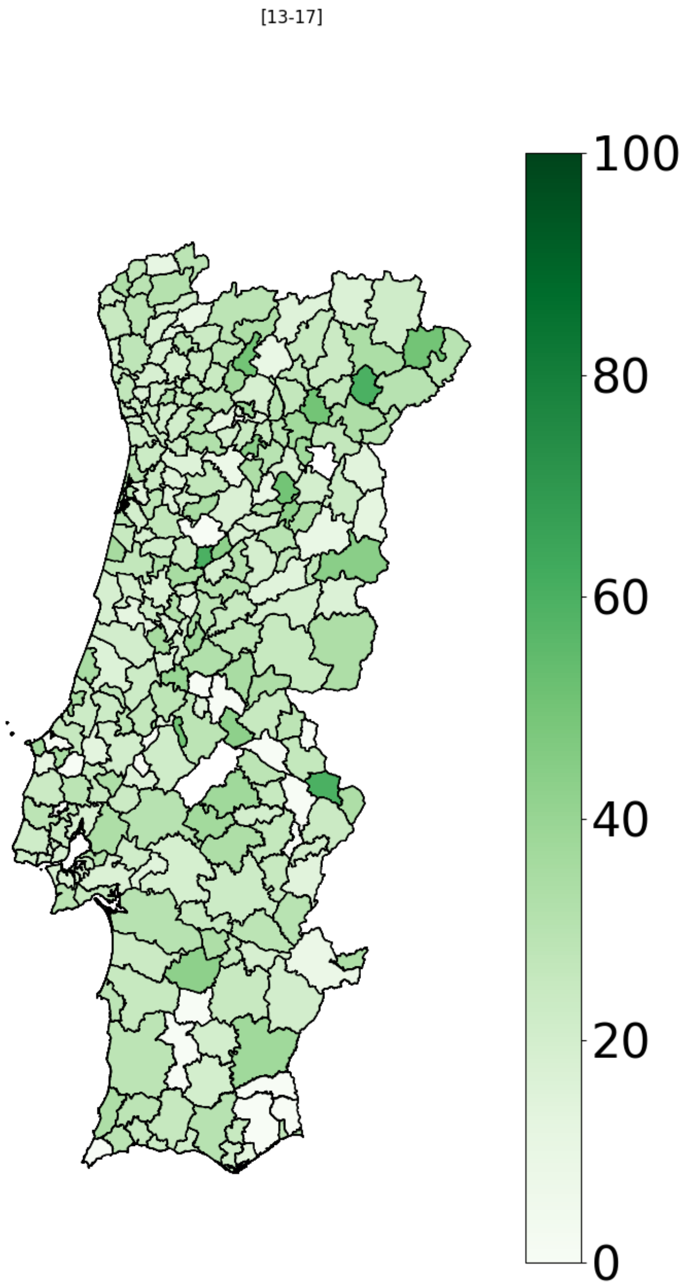


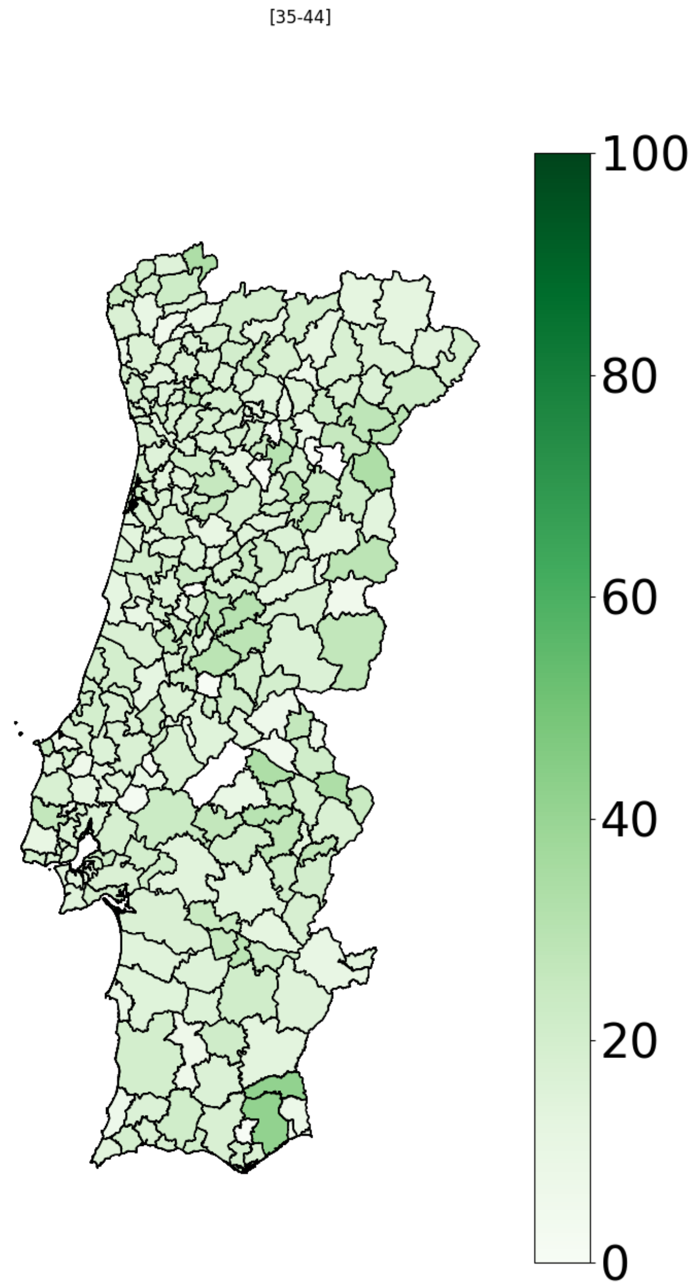
Male

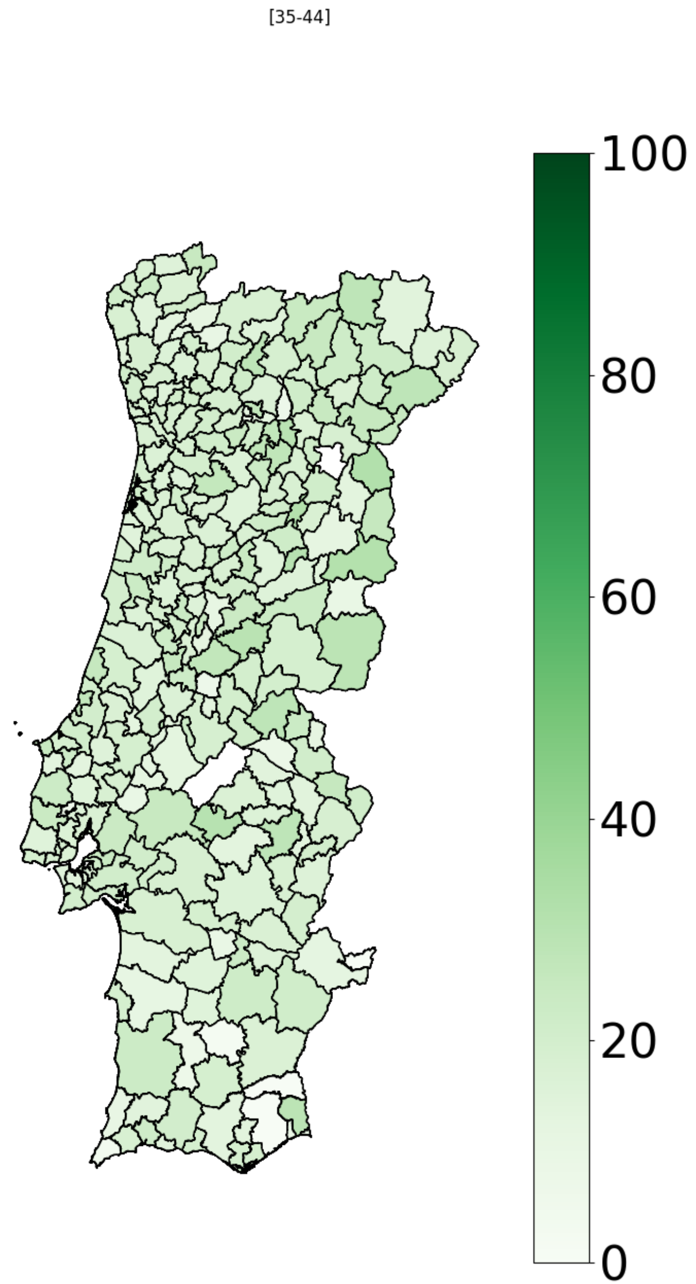


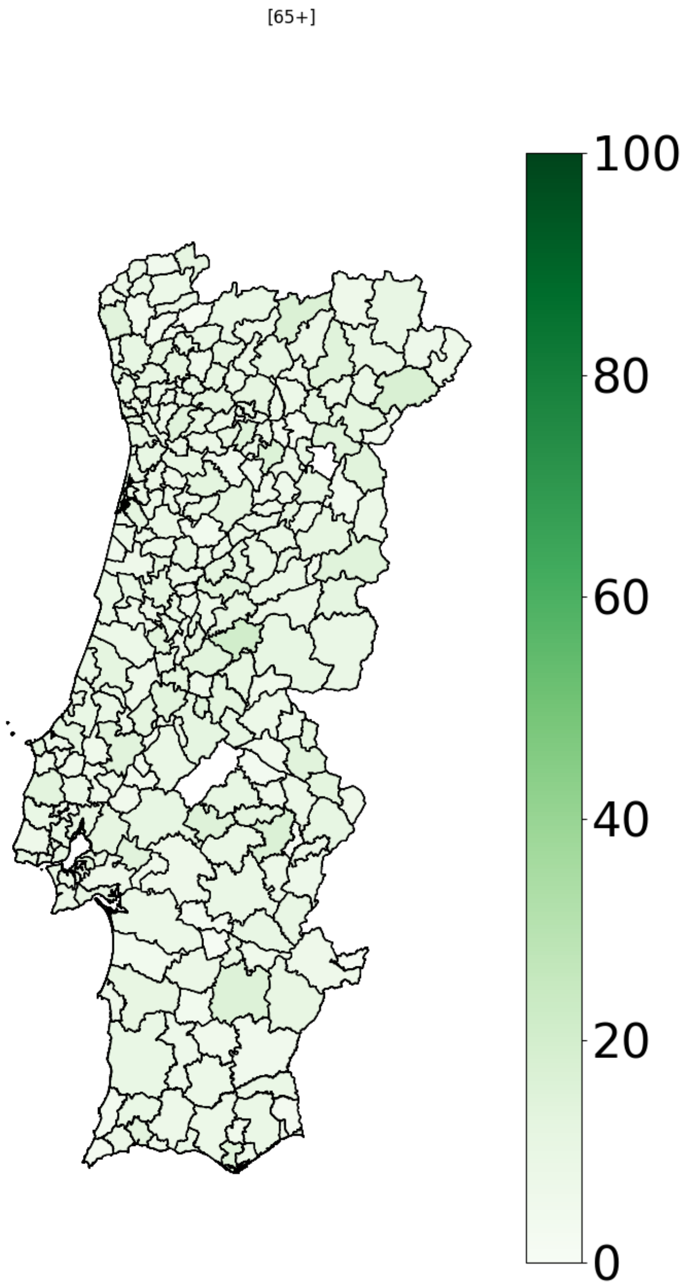
Female

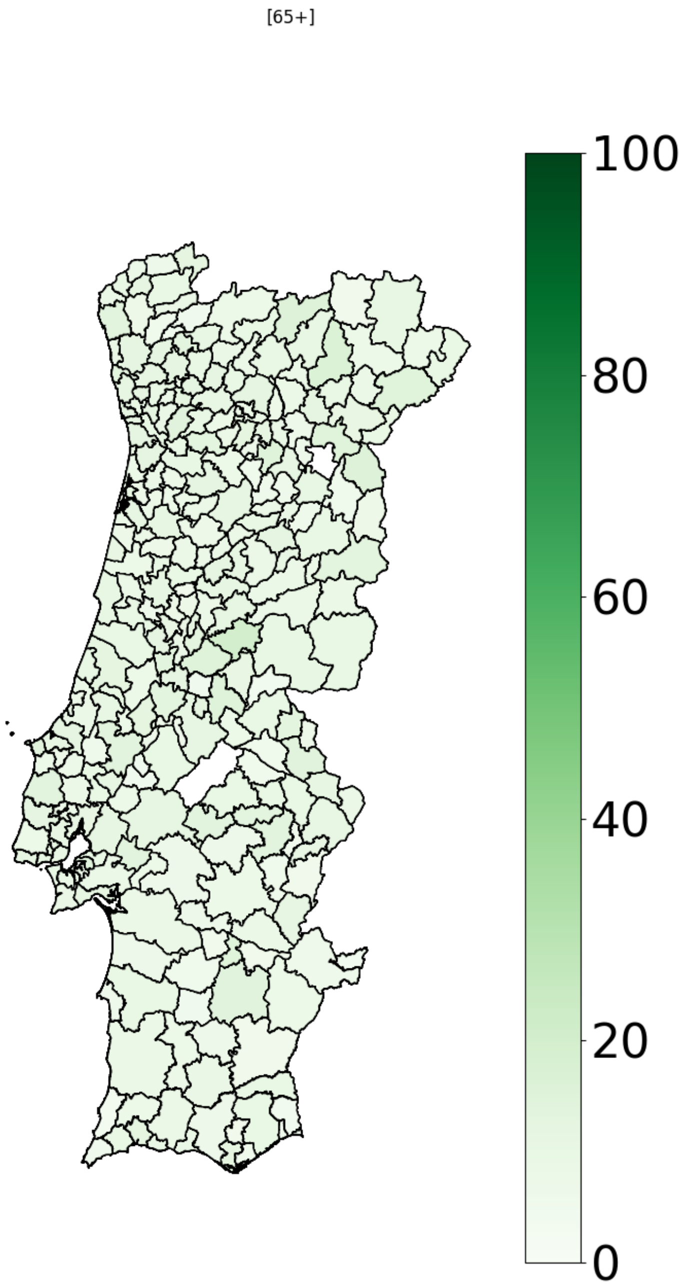

[13-17]
100
80
60
40
20
0


[13-17]
100
80
60
40
20
0


[35-44]
100
80
60
40
20
0

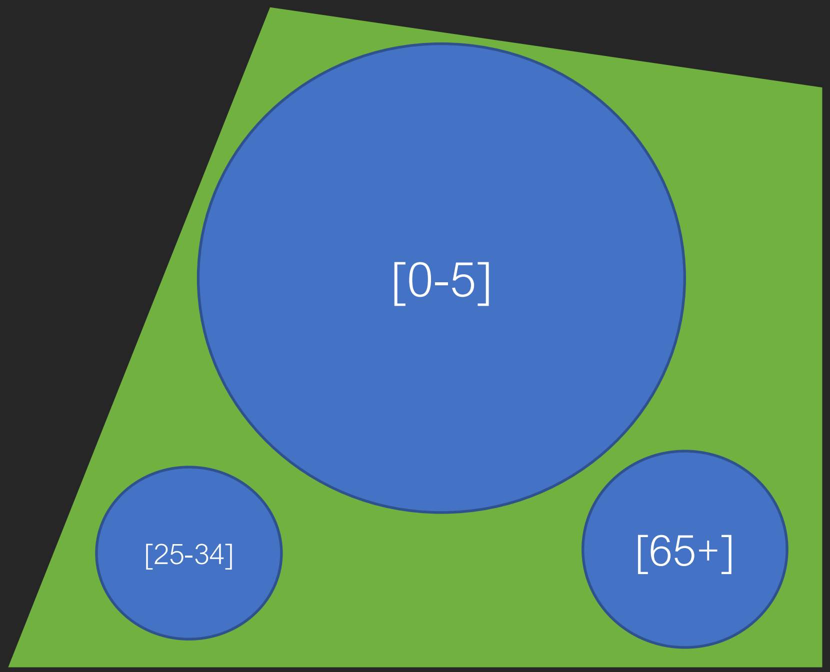

[35-44]
100
80
60
40
20
0


[65+]
100
80
60
40
20
0

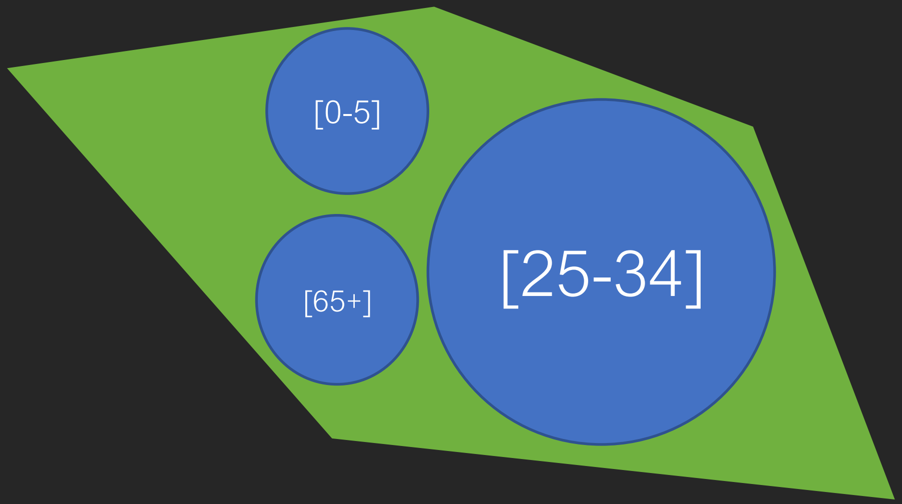

[65+]
100
80
60
40
20
0

Metrics #3:

The final goal would be to standardize the metrics…



Municipality A



Municipality B

We would still need to normalize it!

VI.

Discussion

To calculate the standard population in the usual way we would take the number of people in every municipality by age group and gender…

…the issue was we discovered the number of patients outsmarts the Census population

WHY? Possible reasons:

Random errors when inserting the data in the platforms?....

Random errors when dealing with the data?....

Patients that belong to more than one parameter (two genders, two age groups,…)

Patients that would go to different municipality hospitals?

Although the existence of this issue, we could still calculate a reasonable metrics with the number of visits, for example.

What do the data we have tell us?

- Younger people take more antibiotics
- To be able to do more comparisons, we would dig more the data

The End

Index