"Kluster It! - Integrating K-Means Clustering and Club Convergence Analysis with Shiny"
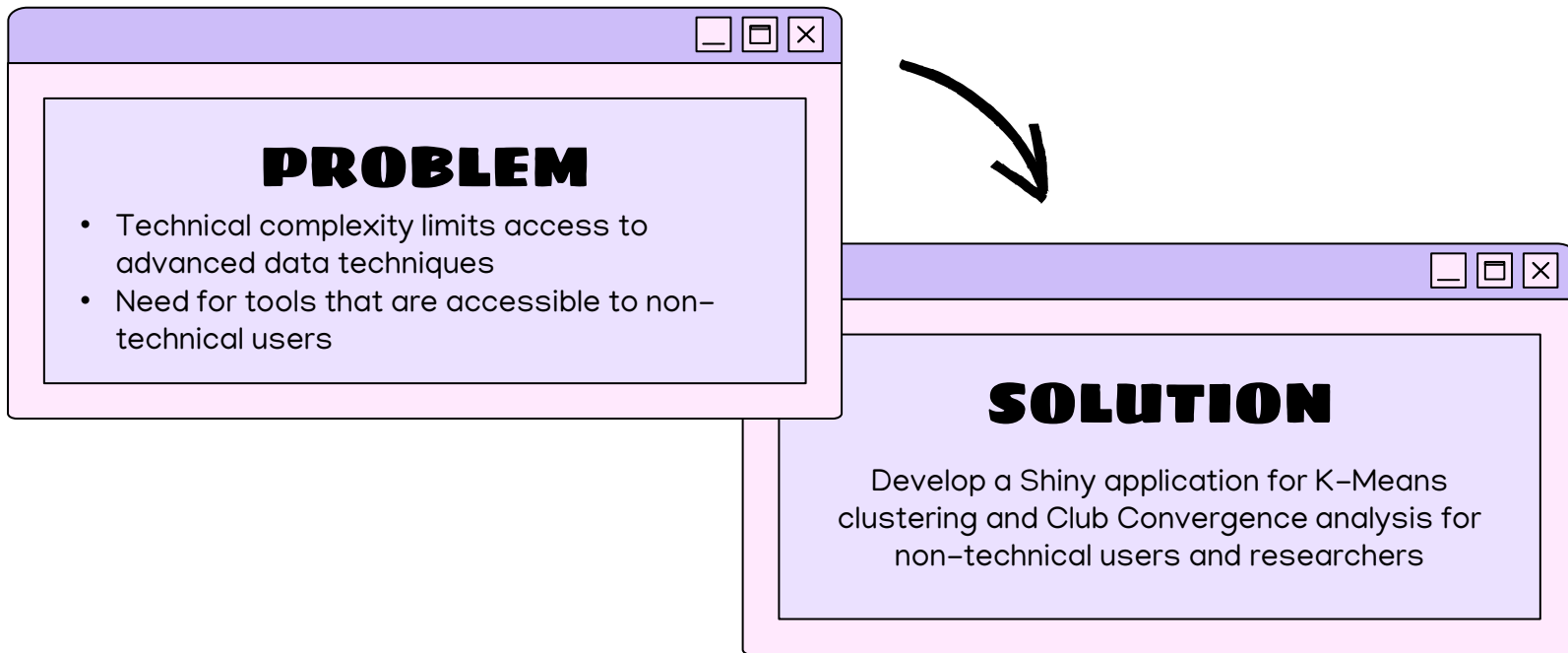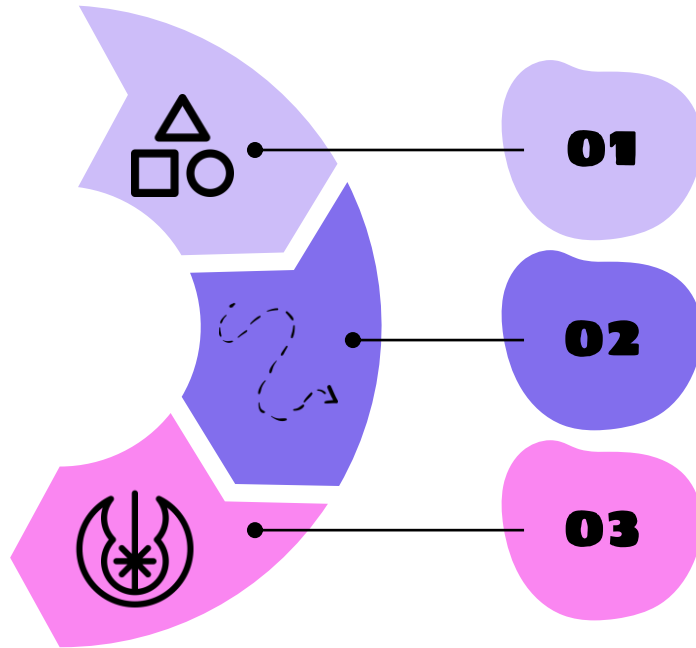
Bernard Banyuls

Kluster It! - Shiny App

PROBLEM
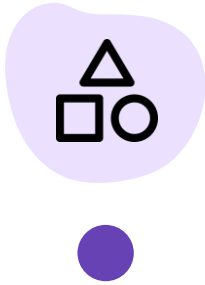
- Technical complexity limits access to advanced data techniques
- Need for tools that are accessible to non-technical users

SOLUTION

Develop a Shiny application for K-Means clustering and Club Convergence analysis for non-technical users and researchers

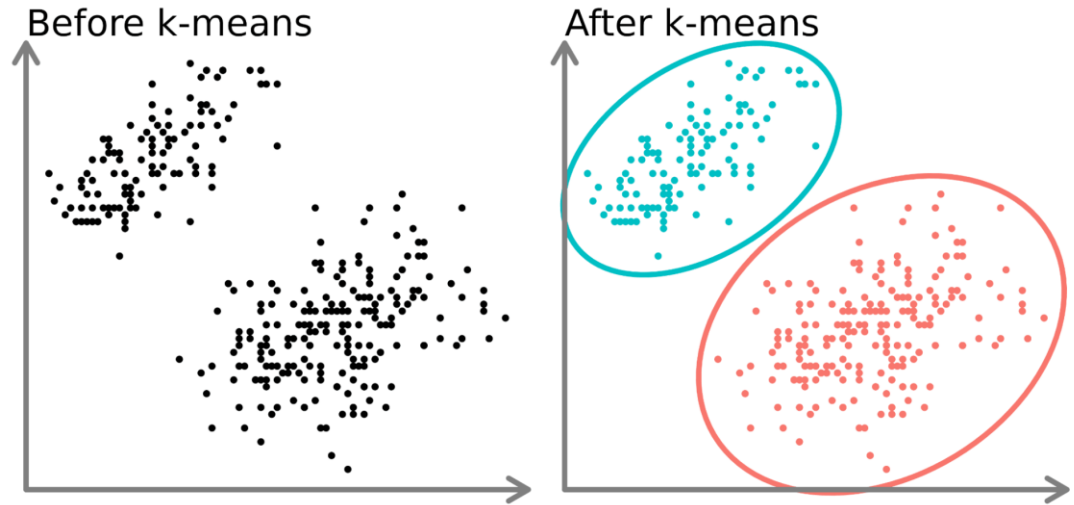# Methodologies

01   KMEANS

02   CLUB CONVERGENCE

03   KLUB INDEX

KMEANS

Partitioning data into clusters based on characteristics

Before k-means

After k-means

CLUB CONVERGENCE
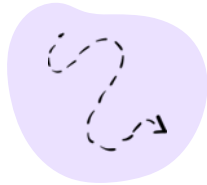
Identifying groups with similar long–term behavior

KLUB INDEX

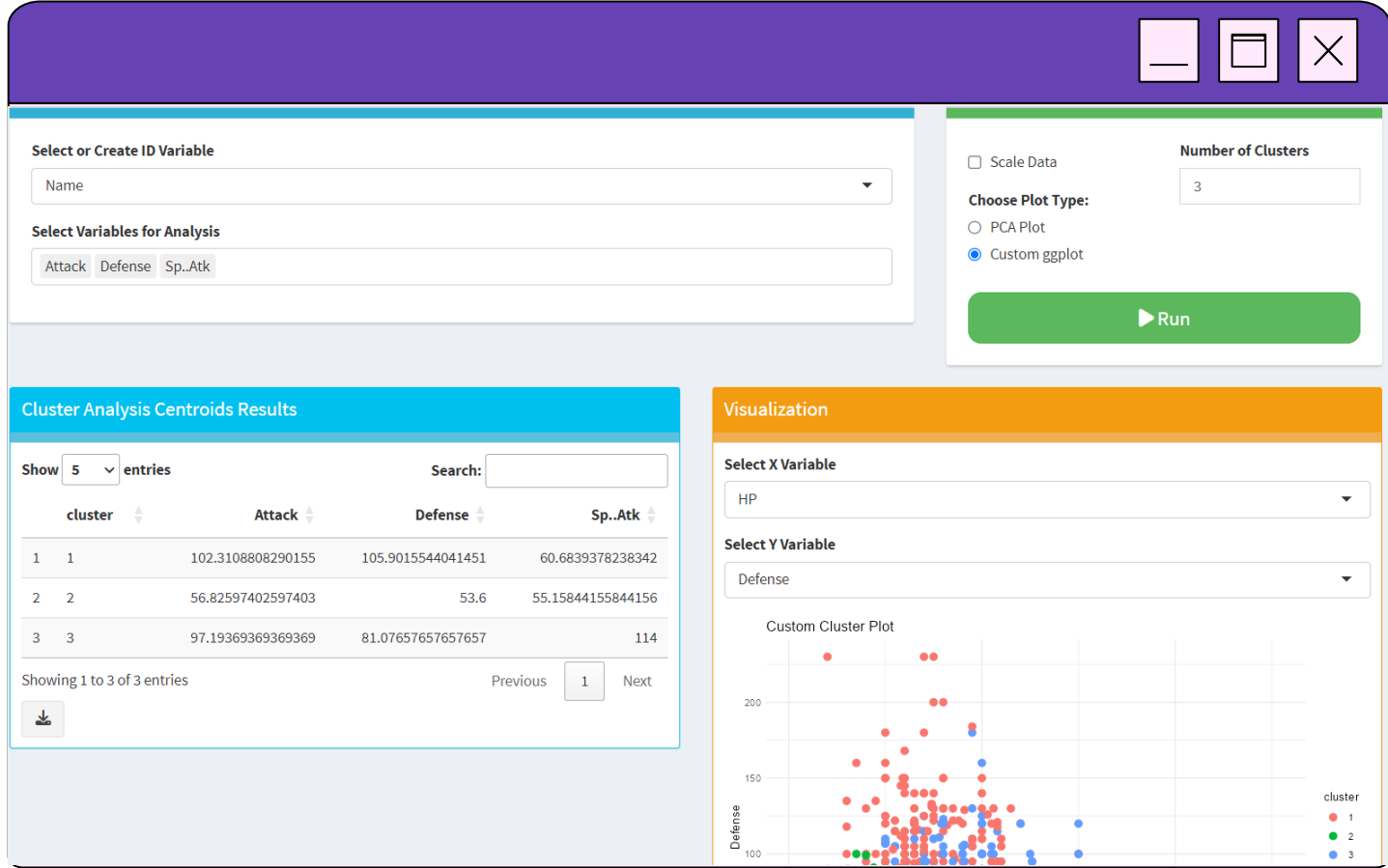| STEP 1 | STEP 2 | STEP 3 | STEP 4 |
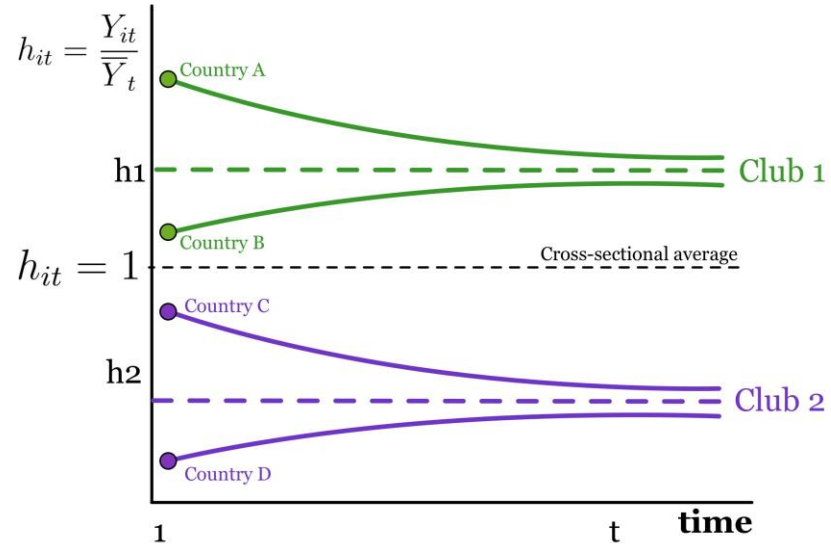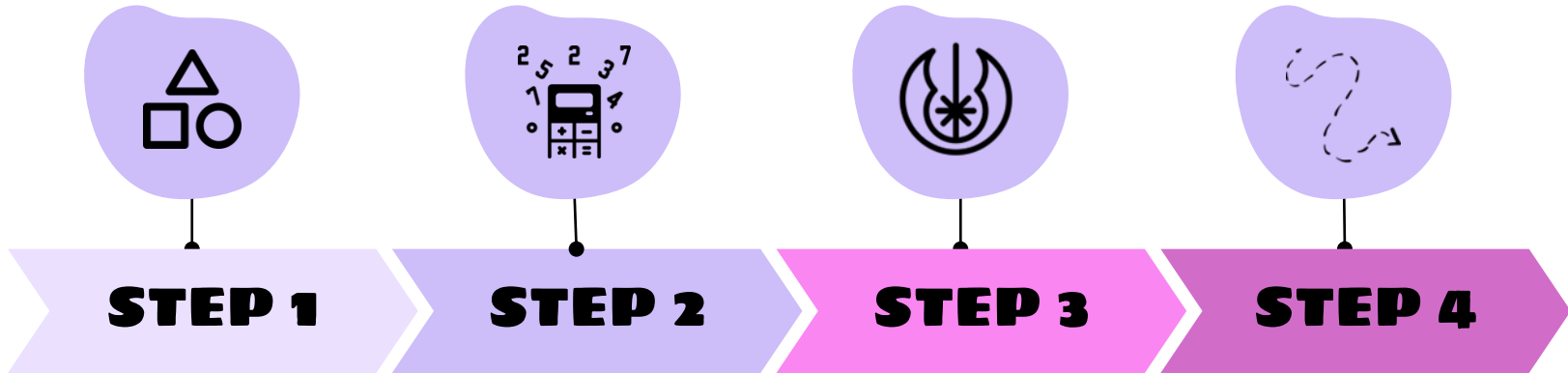
Perform kmeans on entire data to obtain cluster tags.

Use a classification model (XGBoost) to obtain importance coefficients.

Create Index:

$$\text{Index}_{\text{Year}} = \sum_{i=1}^{X} \text{Var}_i^{Year} \times C_i$$

Perform club convergence with the index data.

Select or Create ID Variable

Name ▾

Select Variables for Analysis

Attack Defense Sp..Atk

☐ Scale Data

Number of Clusters

3

Choose Plot Type:

○ PCA Plot

● Custom ggplot

▶ **Run**

Cluster Analysis Centroids Results

Show 5 entries   Search:

	cluster ⇅	Attack ⇅	Defense ⇅	Sp..Atk ⇅
1	1	102.3108808290155	105.9015544041451	60.6839378238342
2	2	56.82597402597403	53.6	55.15844155844156
3	3	97.19369369369369	81.07657657657657	114

Showing 1 to 3 of 3 entries   Previous 1 Next

Visualization

Select X Variable

HP ▾

Select Y Variable

Defense ▾

Custom Cluster Plot

cluster
● 1
● 2
● 3

Select or Create ID Variable

Name ▼

Select Variables for Analysis

Attack Defense Sp..Atk

☐ Scale Data

Number of Clusters

3

Choose Plot Type:

○ PCA Plot

● Custom ggplot

▶ Run

Cluster Analysis Centroids Results

Show [5] entries

Search: [ ]

	cluster	Attack	Defense	Sp..Atk
1	1	102.3108808290155	105.9015544041451	60.6839378238342
2	2	56.82597402597403	53.6	55.15844155844156
3	3	97.19369369369369	81.07657657657657	114

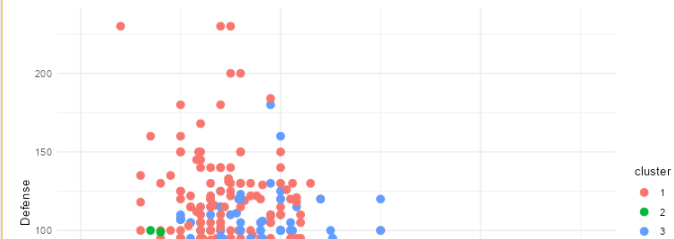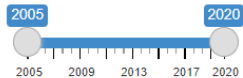Showing 1 to 3 of 3 entries

Previous [1] Next

Visualization

Select X Variable

HP ▼

Select Y Variable

Defense ▼

Custom Cluster Plot

cluster
● 1
● 2
● 3

Select Year Range:

2005 — 2020

2005 2009 2013 2017 2020

Choose View:

- ⦿ All Clubs
- ○ Specific Club

	beta	std.err	tvalue	pvalue
	-0.335	0.045	-7.412	0.000

```
==========================================================================
club 1
--------------------------------------------------------------------------
Denmark, Sweden, Austria, Germany, Belgium, Finland, Luxembourg, Netherlands

beta:      0.2727
std.err:   0.0507
tvalue:    5.3766
pvalue:    1
cstar:     0
```
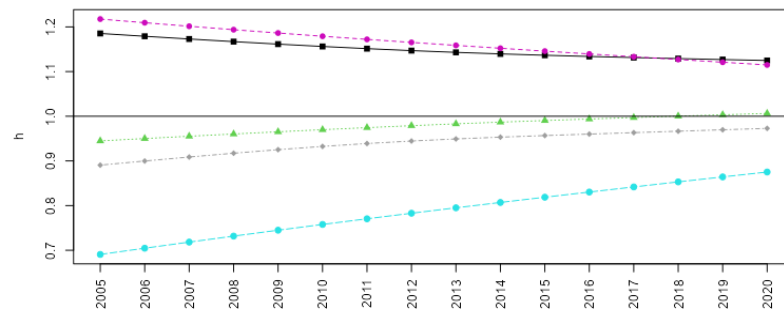

Club 1
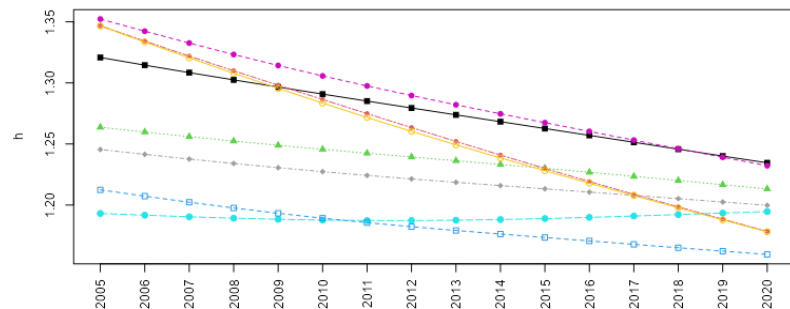
Club 2

Enter Prefix:
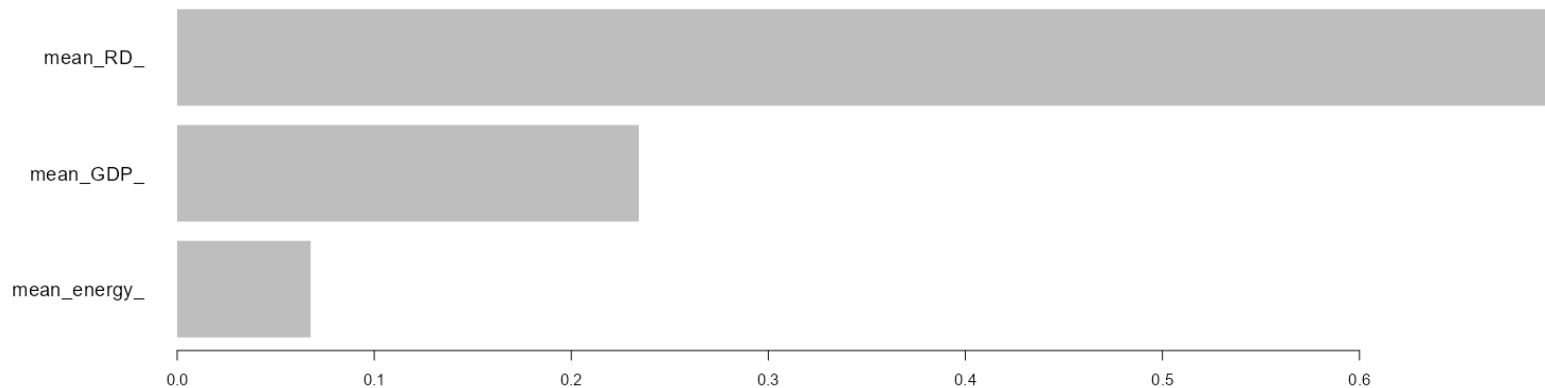
geo ▼

Add Prefix

RD_, GDP_, energy_

▶

Remove All Prefixes

Importance Plot

mean_RD_

mean_GDP_

mean_energy_

0.0   0.1   0.2   0.3   0.4   0.5   0.6

Select ID Variable:
geo ▼

Prefix:
[   ]

[Add Prefix] [Remove Prefix]

View:
- ◉ All Clubs
- ○ Specific

Result:
- ◉ Plot
- ○ Print

Start Year:
2007

End Year:
2016

Coefficient for RD_
0.7

Ranking Order for RD_
- ◉ Ascending
- ○ Descending

Coefficient for GDP_
0.2

Ranking Order for GDP_
- ◉ Ascending
- ○ Descending

Coefficient for energy_
0.1

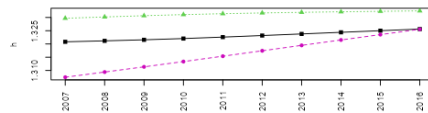## Ranking Order for energy_
- ◉ Ascending
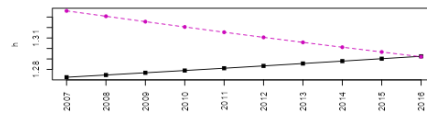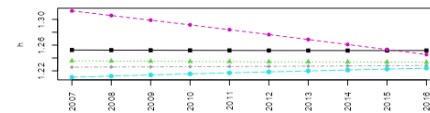- ○ Descending

▶ Run Index Club Analysis

Club 1   Club 2   Club 3

Preview Data —

Show [5] entries   Search: [   ]

	X.	Name	Type.1	Type.2	Total	HP	Attack	Defe
1	1	Bulbasaur	Grass	Poison	318	45	49	
2	2	Ivysaur	Grass	Poison	405	60	62	
3	3	Venusaur	Grass	Poison	525	80	82	
4	3	VenusaurMega Venusaur	Grass	Poison	625	80	100	
5	4	Charmander	Fire		309	39	52	

Showing 1 to 5 of 800 entries

Previous [1] 2 3 4 5 … 160 Next

Variable Types —

	Type
X.	integer
Name	character
Type.1	character
Type.2	character
Total	integer
HP	integer
Attack	integer
Defense	integer
Sp..Atk	integer
Sp..Def	integer
Speed	integer

Modify Data Types —

X.
Select Type
[integer ▾]

Name
Select Type
[character ▾]

Type.1
Select Type
[character ▾]

Type.2
Select Type
[character ▾]

Total
Select Type
[integer ▾]

HP
Select Type
[integer ▾]

Preview Data

Show [5] entries   Search: [_____]

	Country	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1	Belgium	531.5	563.8	600.6	638.7	644	690.7	742.8	795.5	822.1	854.3	900.4	959.5	1045.5	1154.4	1319	1378.8
2	Bulgaria	13.8	15.9	18.4	22.2	24.7	29	29.8	34.6	36.6	46.9	60.4	52.5	54.7	60.1	73.2	75.3
3	Czechia	125.6	149.3	175.6	193.3	184.6	200.3	243.4	273.9	285	294	308.4	280.8	324.5	377.6	408.3	400.8
4	Denmark	941.3	998.5	1077.7	1223.8	1282	1281.6	1312.7	1360	1371.8	1376.2	1473.7	1534.2	1503.2	1551.1	1568.6	1624.8
5	Germany	677.3	715.3	747.1	810	818	855.9	942	984.8	990.1	1043.1	1093.4	1121.7	1206.4	1264.2	1325.3	1273.2

Showing 1 to 5 of 26 entries

Previous [1] 2 3 4 5 6 Next

Reorder Columns

| Country | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
| 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |

[Apply Reorder]

Rename Columns

Country	**2005**	**2006**	**2007**	**2008**	**2009**
Country	2005	2006	2007	2008	2009

2010	**2011**	**2012**	**2013**	**2014**	**2015**
2010	2011	2012	2013	2014	2015

2016	**2017**	**2018**	**2019**	**2020**
2016	2017	2018	2019	2020

[Apply Rename]