# World Development Indicators

*Which country will develop more?*


kaggle

Andrea Schiavon
Leonardo Comandini
Stefano Moawad

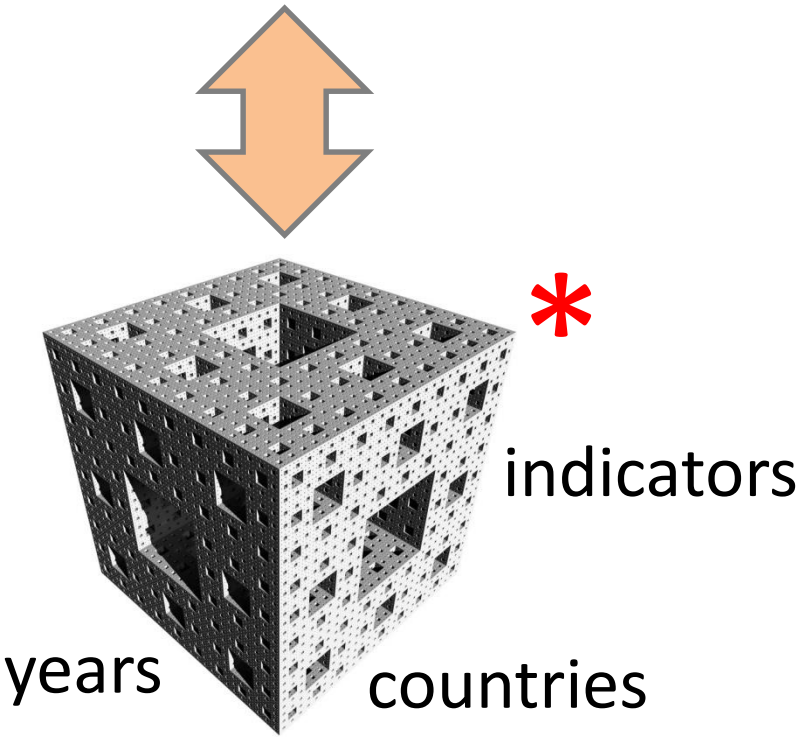
# Overview

i. Results with Shiny

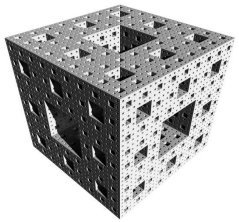
ii. Growth Analysis

iii. Cluster Analysis

| | CountryName | CountryCode | IndicatorName | IndicatorCode | Year | Value |
|---|---|---|---|---|---|---|
| 1 | Arab World | ARB | Adolescent fertility rate (births per 1,000 women age… | SP.ADO.TFRT | 1960 | 1.335609e+02 |
| 2 | Arab World | ARB | Age dependency ratio (% of working-age population) | SP.POP.DPND | 1960 | 8.779760e+01 |
| 3 | Arab World | ARB | Age dependency ratio, old (% of working-age populati… | SP.POP.DPND.OL | 1960 | 6.634579e+00 |
| 4 | Arab World | ARB | Age dependency ratio, young (% of working-age pop… | SP.POP.DPND.YG | 1960 | 8.102333e+01 |
| 5 | Arab World | ARB | Arms exports (SIPRI trend indicator values) | MS.MIL.XPRT.KD | 1960 | 3.000000e+06 |
| 6 | Arab World | ARB | Arms imports (SIPRI trend indicator values) | MS.MIL.MPRT.KD | 1960 | 5.380000e+08 |
| 7 | Arab World | ARB | Birth rate, crude (per 1,000 people) | SP.DYN.CBRT.IN | 1960 | 4.769789e+01 |
| 8 | Arab World | ARB | CO2 emissions (kt) | EN.ATM.CO2E.KT | 1960 | 5.956399e+04 |
| ⋮ | | | ⋮ | ⋮ | | |
| 5656455 | Zimbabwe | ZWE | Time required to start a business (days) | IC.REG.DURS | IC.PRP.DURS | 9.000000e+01 |
| 5656456 | Zimbabwe | ZWE | Time to prepare and pay taxes (hours) | IC.TAX.DURS | 2015 | 2.420000e+02 |
| 5656457 | Zimbabwe | ZWE | Time to resolve insolvency (years) | IC.ISV.DURS | 2015 | 3.300000e+00 |
| 5656458 | Zimbabwe | ZWE | Total tax rate (% of commercial profits) | IC.TAX.TOTL.CP.ZS | 2015 | 3.280000e+01 |



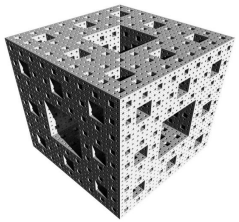

*much more irregular



Dataset

Goal

*Which countries will develop more?*



Dataset

Extract the most interesting features for the main topics

Goal

*Which countries will develop more?*

# How we show the results

## Shiny

by RStudio

**A web application framework for R**

Turn your analyses into interactive web applications

No HTML, CSS, or JavaScript knowledge required

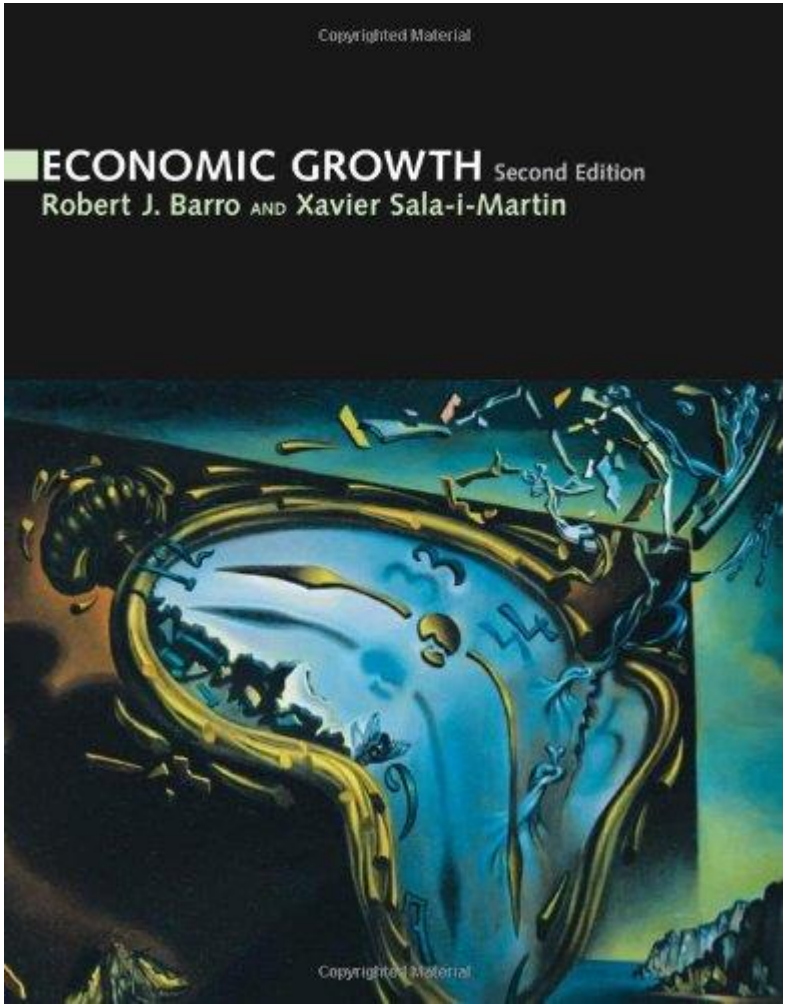
TUTORIAL ARTICLES GALLERY REFERENCE DEPLOY HELP

Get inspired
(gallery)

Get started
(tutorial)

Go deeper
(articles)

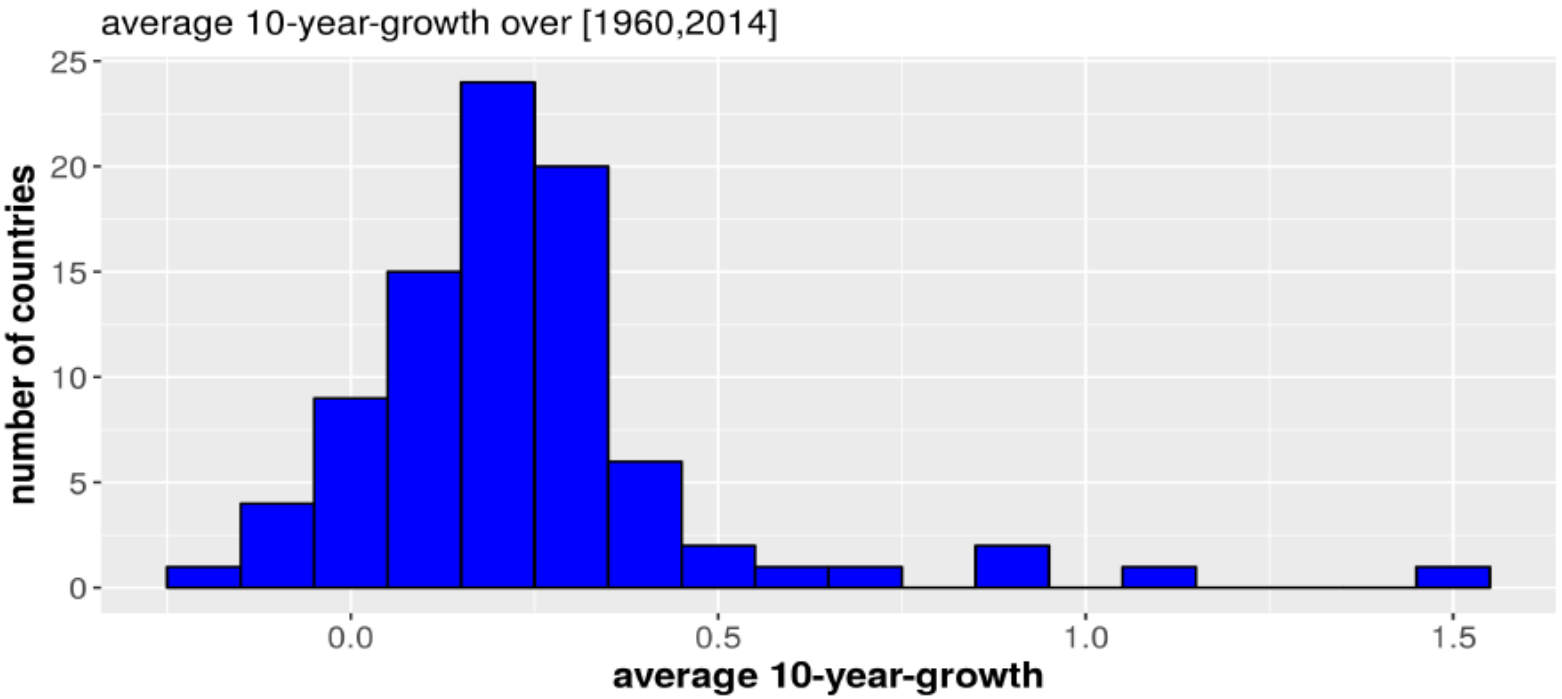
# Growth




Empirical evidences

Explanatory model for 10-year-Growth

Prediction, Evaluation and Comparison

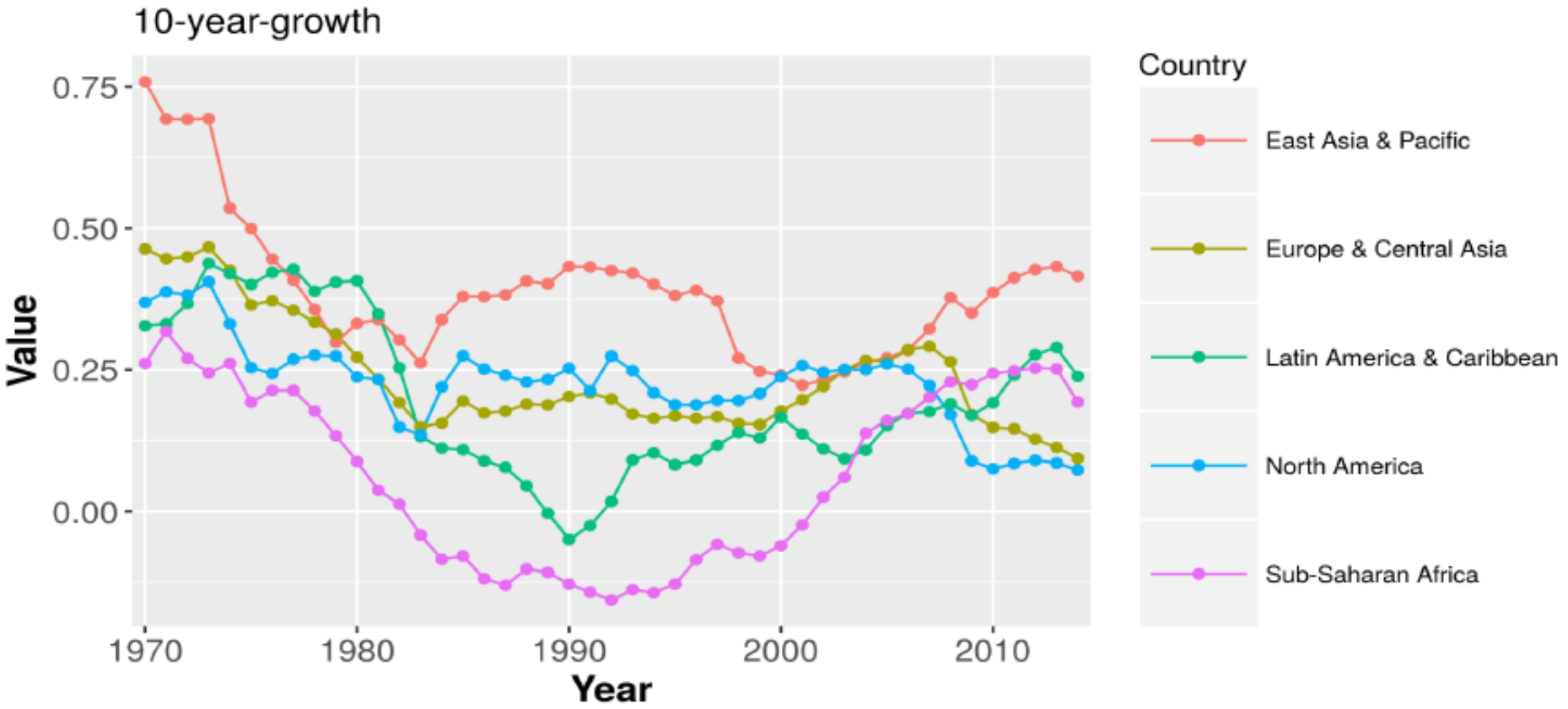



The 10-year-Growth is the 10-year percentage variation of the GDP per capita in local currency. More formally,

$$Growth_t := \frac{GDP_t - GDP_{t-10}}{GDP_{t-10}} \quad (1)$$

where $GDP$ is the Gross Domestic Product per capita

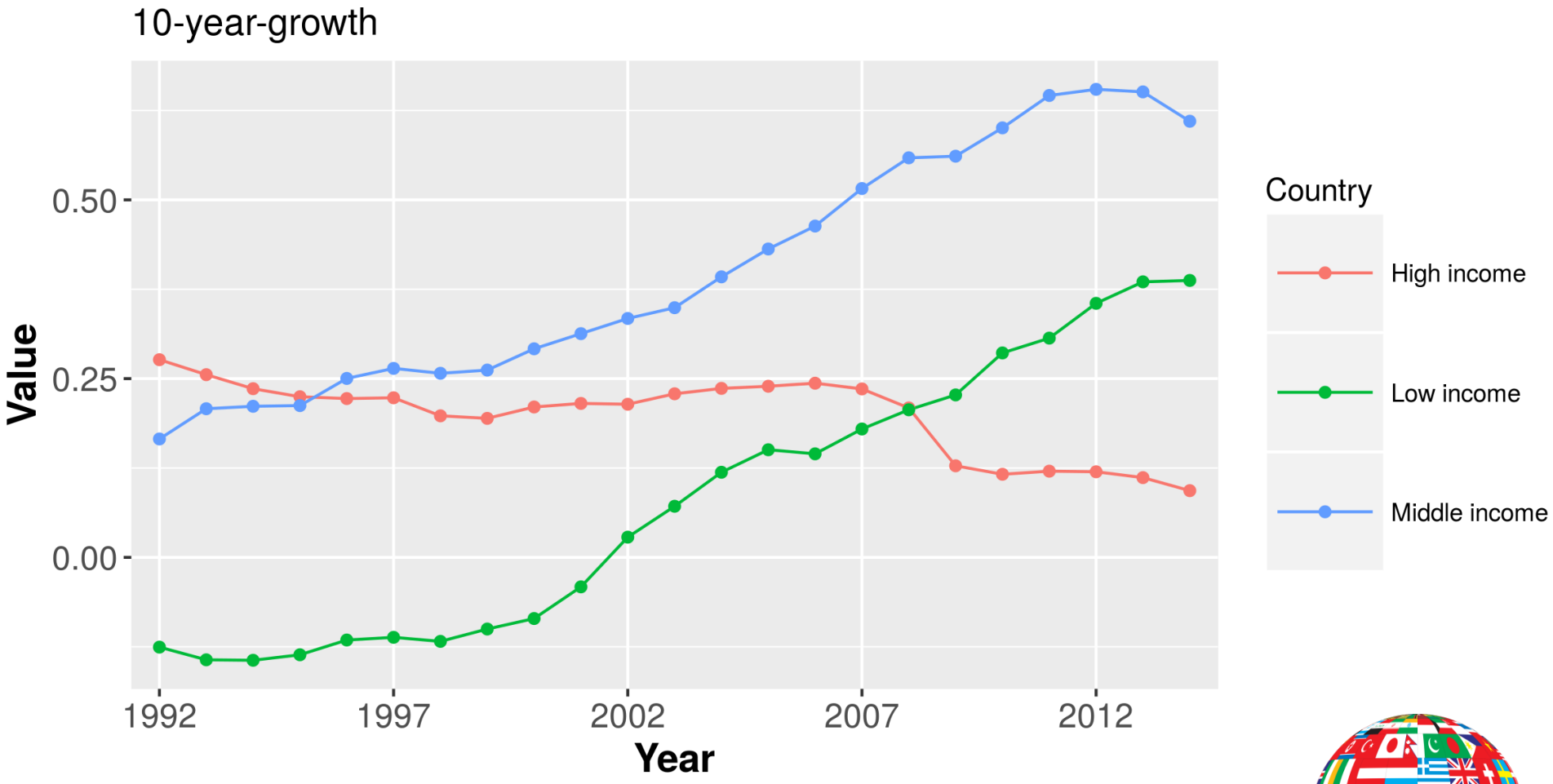
# 10-year-growth by region




Differences between decades → Dummy for decades

Differences between regions → Dummy for Asia and Africa

# 10-year-growth by Income group




Differences between Income group

Dummy for High Income and Low Income

# The Regressors: State and Environmental variables

Education := $\frac{\text{tot enrolment primary school}}{\text{population}}$ [%]

Health := $\frac{1}{\text{life expectancy at birth}}$ $[\text{year}]^{-1}$

Fertility := average number of births per woman

State variables $\underline{h}_t$

Inflation [%]

GDP := log(GDP)

FDI := financial capital owned by foreign investors [% of GDP]

Openess := $\frac{\text{Inport}+\text{Export}}{\text{GDP}}$

Consumption := households consumption expenditure [% of GDP]

Investment := government expenditures for goods and services [% of GDP]

Enviromental variables $\underline{y}_t$

$$\Longrightarrow \; Growth_t = F(\underline{h}_{t-10}, \underline{y}_{t-10})$$

# Complete model

Let $\epsilon \sim N(0, \sigma^2)$

$$Growth_{glm} = \beta_{0glm} + \beta_{1glm}\text{fertility} + \beta_2\text{FDI} + \beta_{3glm}\text{GDP} + \beta_4\text{education} + \beta_5\text{consumption} + \beta_6\text{inflation} + \beta_7\text{health} + \beta_{8glm}\text{investment} + \beta_9\text{openess} + \epsilon$$

$$g \in \{[1983, 1993], [1993, 2003], [2003, 2013]\}$$

$$l \in \{\text{Asia}, \text{Africa}, \text{Others}\}$$

$$m \in \{\text{High Income}, \text{Medium Income}, \text{Low Income}\}$$

Stepwise regression → Reduced model

| | | | |
|---|---|---|---|
| (Intercept) | 0.9531 (0.3791)* | fertility:I1 | 0.0804 (0.0297)** |
| fertility | −0.0849 (0.0244)*** | investment:I1 | −0.0354 (0.0073)*** |
| FDI | −0.0085 (0.0063) | investment:I2 | 0.0327 (0.0088)*** |
| GDP | −0.0903 (0.0305)** | GDP:R1 | −0.3070 (0.0348)*** |
| education | −0.0025 (0.0010)* | fertility:R1 | −0.3880 (0.0362)*** |
| consumption | 0.0047 (0.0010)*** | fertility:R2 | −0.0527 (0.0274) |
| health | −21.0428 (11.8102) | investment:R2 | −0.0425 (0.0073)*** |
| R1 | 3.8459 (0.3718)*** | | |
| R2 | 0.8626 (0.1585)*** | $R^2$ | 0.8705 |
| I1 | 1.0546 (0.4503)* | Adj. $R^2$ | 0.8364 |
| I2 | −0.4912 (0.1445)** | Num. obs. | 116 |
| investment | 0.0407 (0.0063)*** | RMSE | 0.1134 |
| D1 | −0.3408 (0.1521)* | | |
| D2 | −0.4913 (0.1369)*** | | |
| GDP:D1 | 0.0841 (0.0182)*** | | |
| investment:D1 | −0.0189 (0.0048)*** | | |
| GDP:D2 | 0.0640 (0.0163)*** | | |
| GDP:I1 | −0.0783 (0.0474) | | |

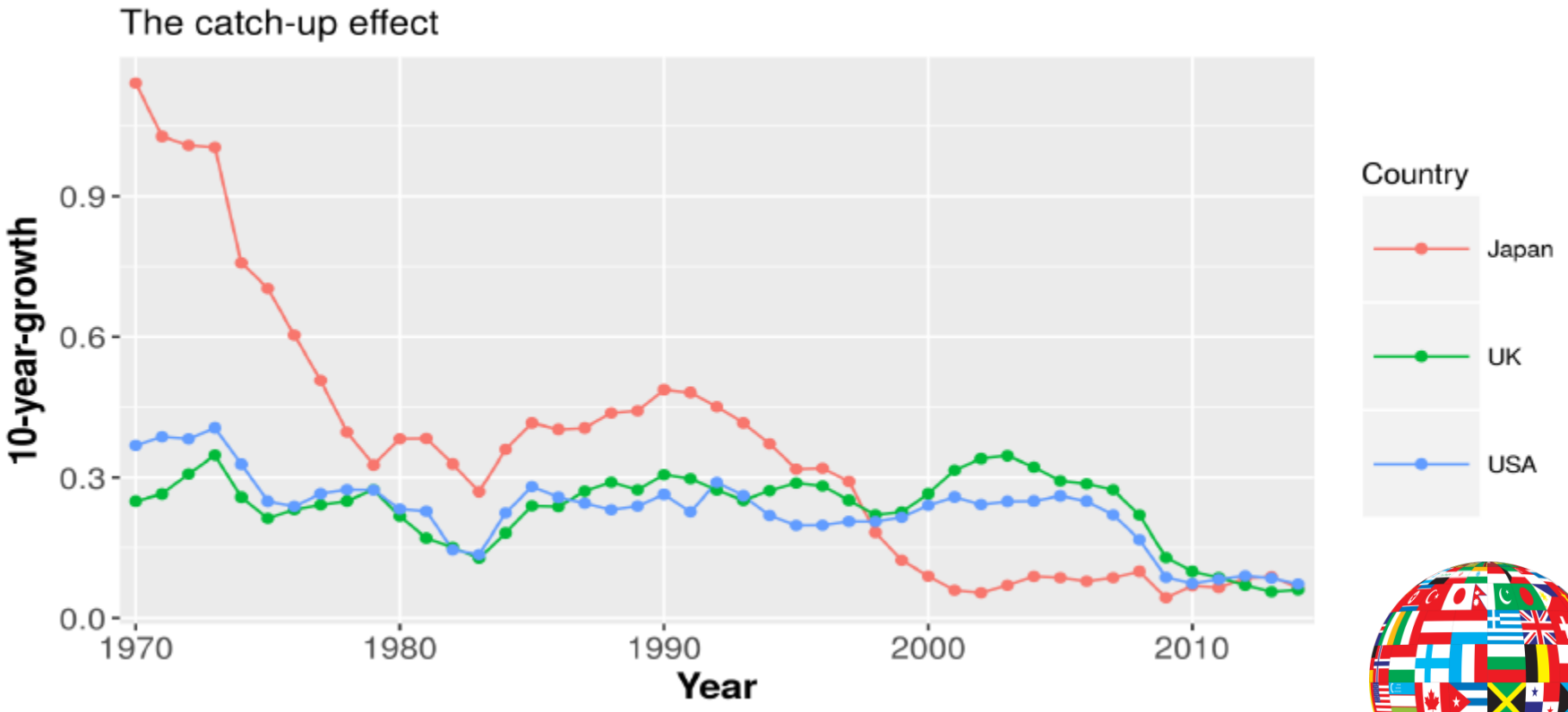
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Legend:

- D1 = [1983,1993) D2 = [1993,2003)
- R1 = Asia R2 = Africa
- I1 = high income
- I2 = low income

# Results (1/2): The Catch-Up Effect

**conditional convergence principle**: the lower the initial GDP the higher the growth over the next decade



# Results (2/2): Fertility

Asia – Middle Income

$\hat{\beta}_1 = -0.4729$

Europe – High Income

$\hat{\beta}_1 = -0.0045$


# Prediction model

Let $\epsilon \sim N(0, \sigma^2)$

$$Growth_{lm} = \beta_{0lm} + \beta_{1lm}\text{fertility} + \beta_2\text{FDI} + \beta_{3lm}\text{GDP} + \beta_4\text{education} + \beta_5\text{consumption} + \beta_6\text{inflation} + \beta_7\text{health} + \beta_{8lm}\text{investment} + \epsilon$$

$$l \in \{\text{Asia}, \ \text{Africa}, \ \text{Others}\}$$

$$m \in \{\text{High Income}, \ \text{Medium Income}, \ \text{Low Income}\}$$

# Predictor evaluation

fitting sample = [1983,2013] test sample = [2003,2013]

$F_t$ = prediction for the growth in $t$ with our model

$Y_t$ = realization of growth in $t$

$e_t$ = prediction error

$ME = \sum_{t=0}^{n} \frac{1}{n} e_t$ = mean error

$MAD = \sum_{t=0}^{n} \frac{1}{n} \|e_t\|$ = mean absolute deviation

$RMSE = \sqrt{\sum_{t=0}^{n} \frac{1}{n} e_t^2}$ = root mean square error

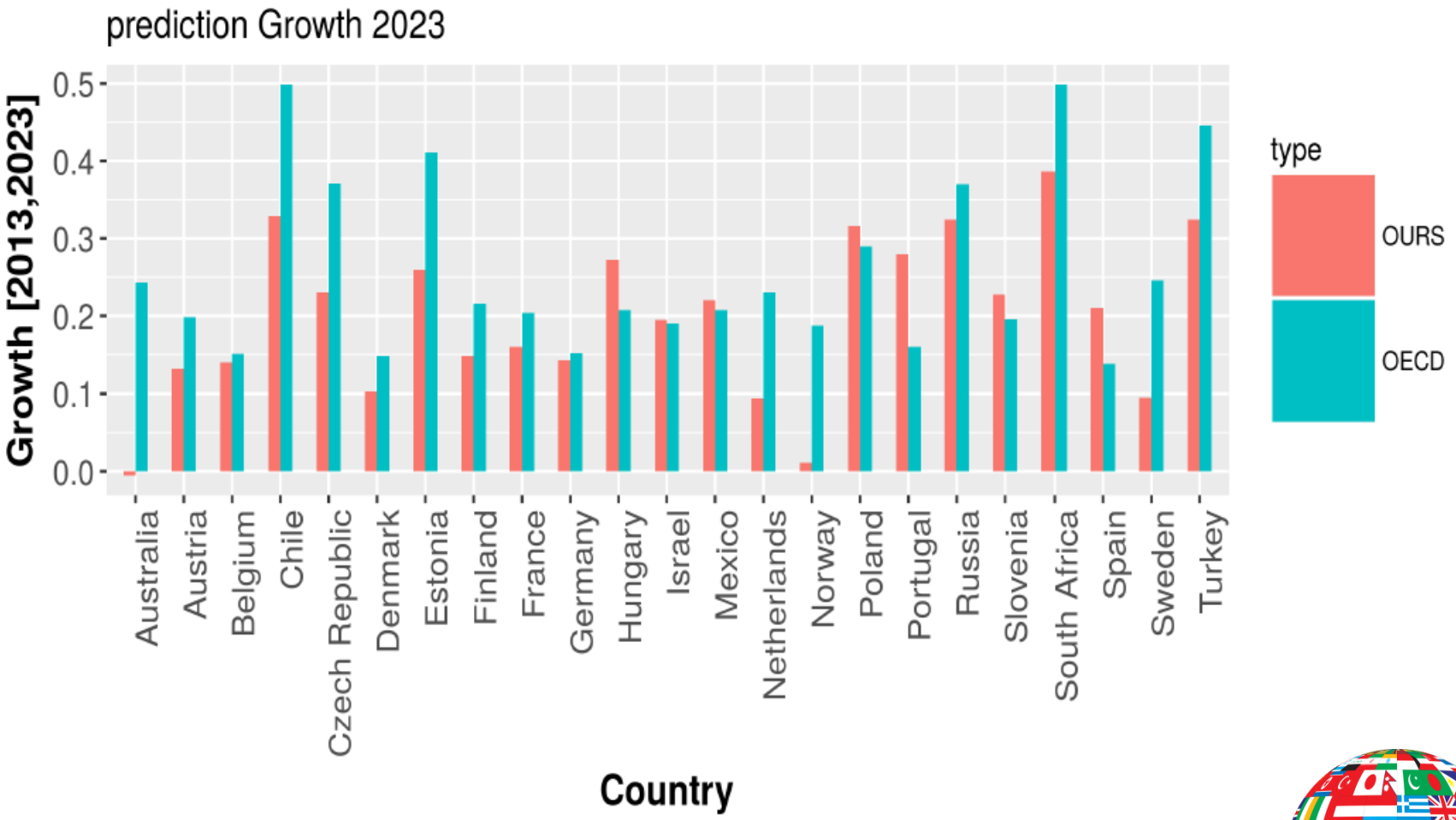
validation on $n = 12$ new countries

| ME | MAD | RMSE |
|---|---|---|
| 0.032 | 0.163 | 0.211 |

slightly overestimating

inaccurate out-of-sample

# Comparison with OECD predictions




**OECD** = The Organisation for Economic Co-operation and Development is an intergovernmental economic organisation with 35 member countries, founded in 1960 to stimulate economic progress and world trade

Steps to clustering

1344 indicators

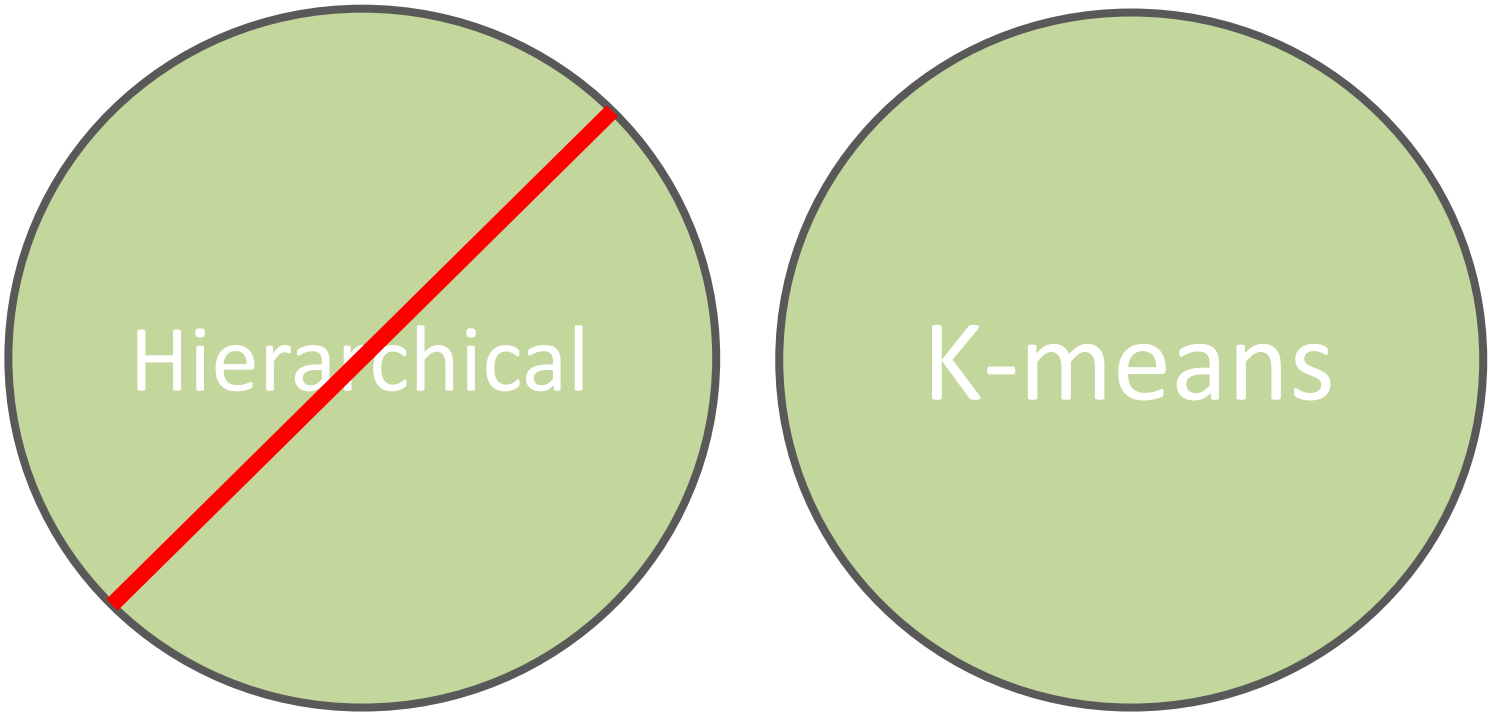
Filtering

Topics

PCA

Clustering

# Clustering

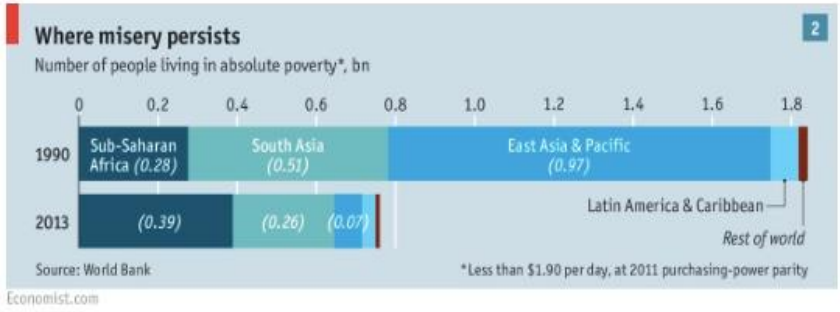



The Economist

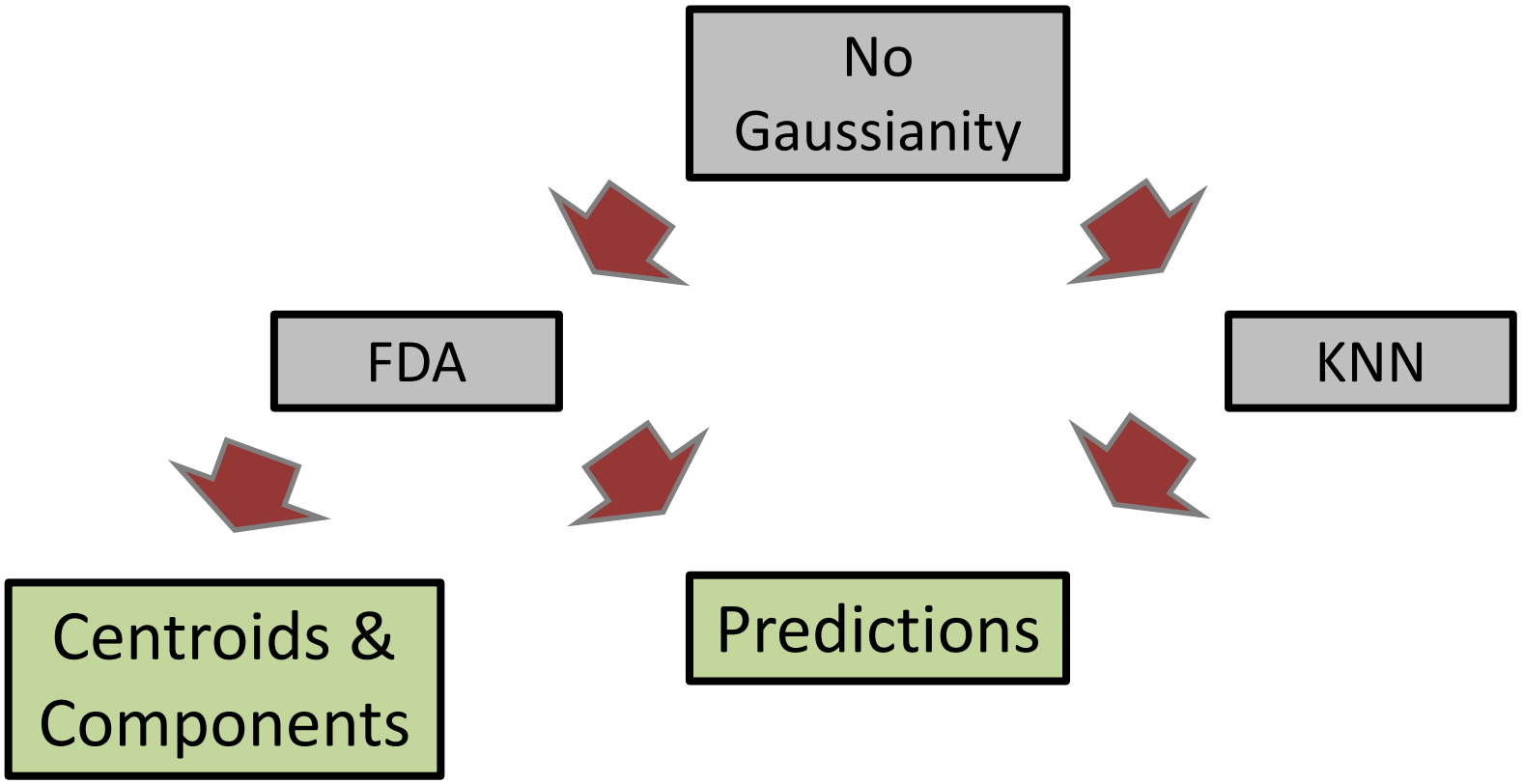
Israel attracts about 15% of the world's venture-capital investment in cyber-security. It is part of Israel's booming "startup-nation" economy, the most dynamic innovation ecosystem outside America.

some of Russia's economic strengths, such as its consistent trade surpluses and its substantial foreign-exchange reserves

massive incubator he is funding for 1,000 startups, opens in Paris soon. But for such companies to scale up fast, as American ones do, he says that Europe needs to "unify all fiscal rules and norms" into a true single market.




# Discrimination Analysis




# FDA or KNN

| | APERCV KNN | APERCV FDA |
|---|---|---|
| Agricolture | 0.1275 | 0.2617 |
| Economic indicators | 0.0712 | 0.1905 |
| Ease to start a business | 0.0643 | 0.1462 |
| Natural resources | 0.0393 | 0.0561 |
| Production | 0.0559 | 0.3230 |
| Telecommunication | 0.0653 | 0.0151 |
| Trade | 0.1361 | 0.0651 |