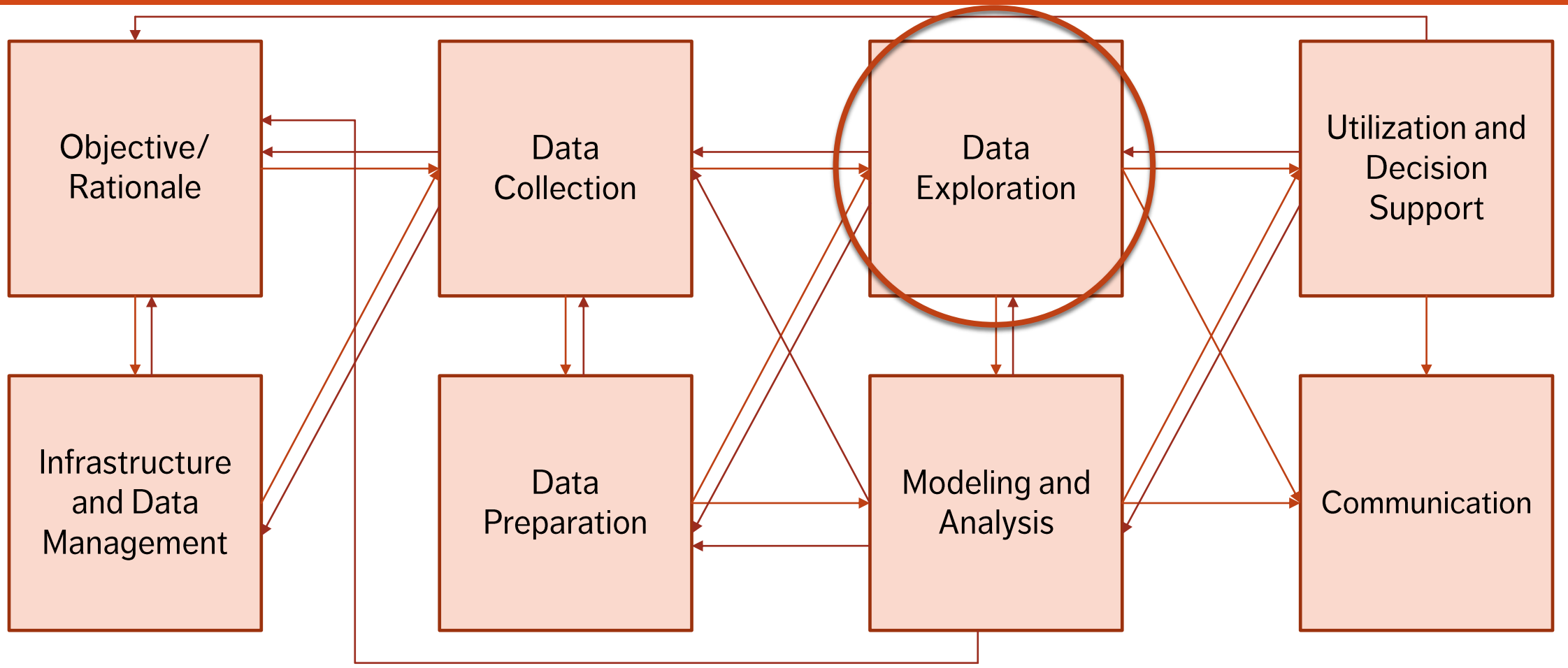
DATA EXPLORATION

THE (MESSY) ANALYSIS PROCESS



NON-VISUALIZATION SUMMARIES

```
autumn spring summer winter
    80     84     86     90
```

```
       Cl              NO3              NH4
 Min.   :  0.222   Min.   : 0.000   Min.   :    5.00
 1st Qu.: 10.994   1st Qu.: 1.147   1st Qu.:   37.86
 Median : 32.470   Median : 2.356   Median :  107.36
 Mean   : 42.517   Mean   : 3.121   Mean   :  471.73
 3rd Qu.: 57.750   3rd Qu.: 4.147   3rd Qu.:  244.90
 Max.   :391.500   Max.   :45.650   Max.   :24064.00
 NA's   :16        NA's   :2        NA's   :2
```

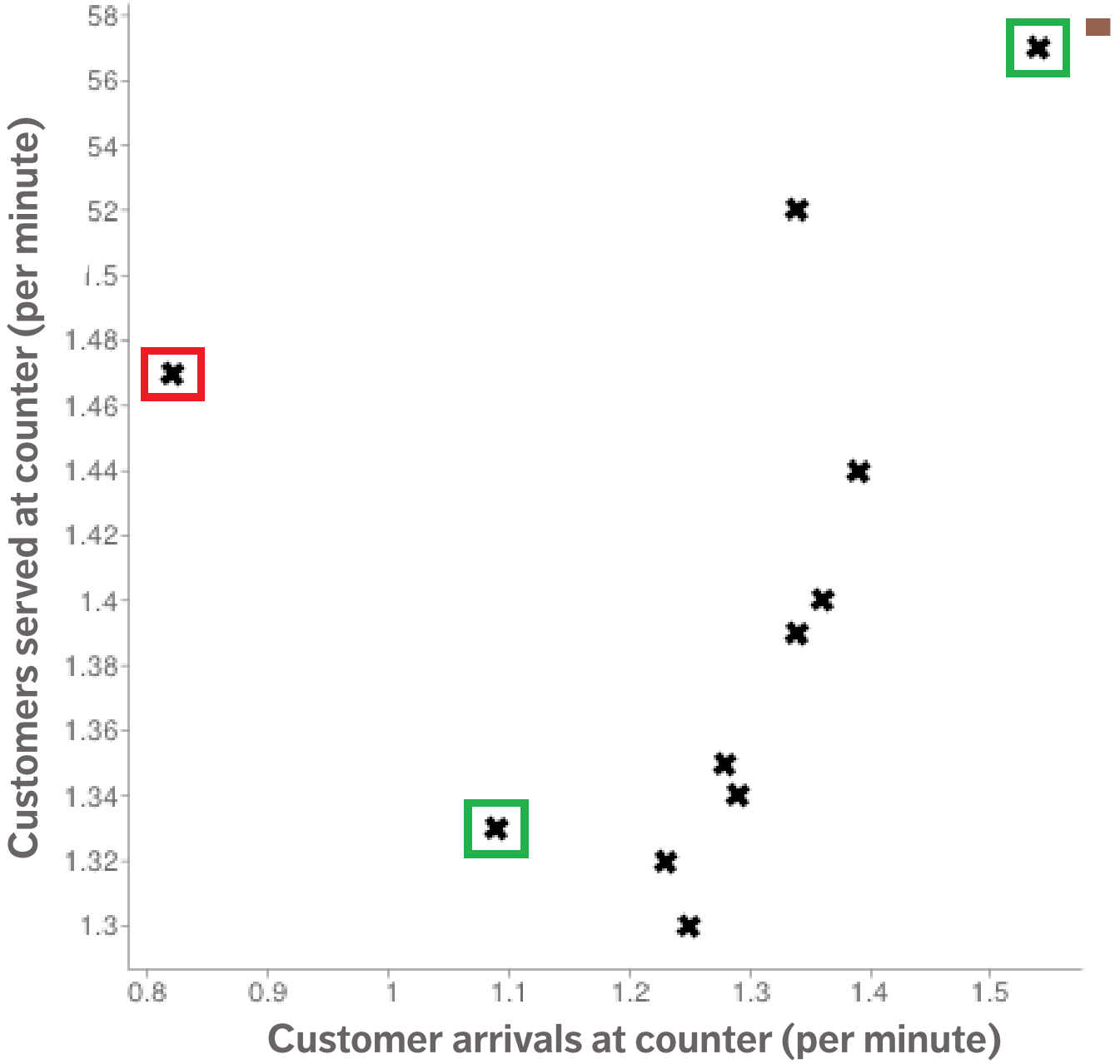
```
    season
 Length:340
 Class :character
 Mode  :character
```

PRE-ANALYSIS USE

Data visualization can be used to set the stage for analysis:

- **detecting anomalous entries**
  invalid entries, missing values, outliers
- **shaping the data transformations**
  binning, standardization, Box-Cox transformations, PCA-like transformations
- **getting a sense for the data**
  data analysis as an art form, exploratory analysis
- **identifying hidden data structure**
  clustering, associations, patterns informing the next stage of analysis

[Personal dataset]

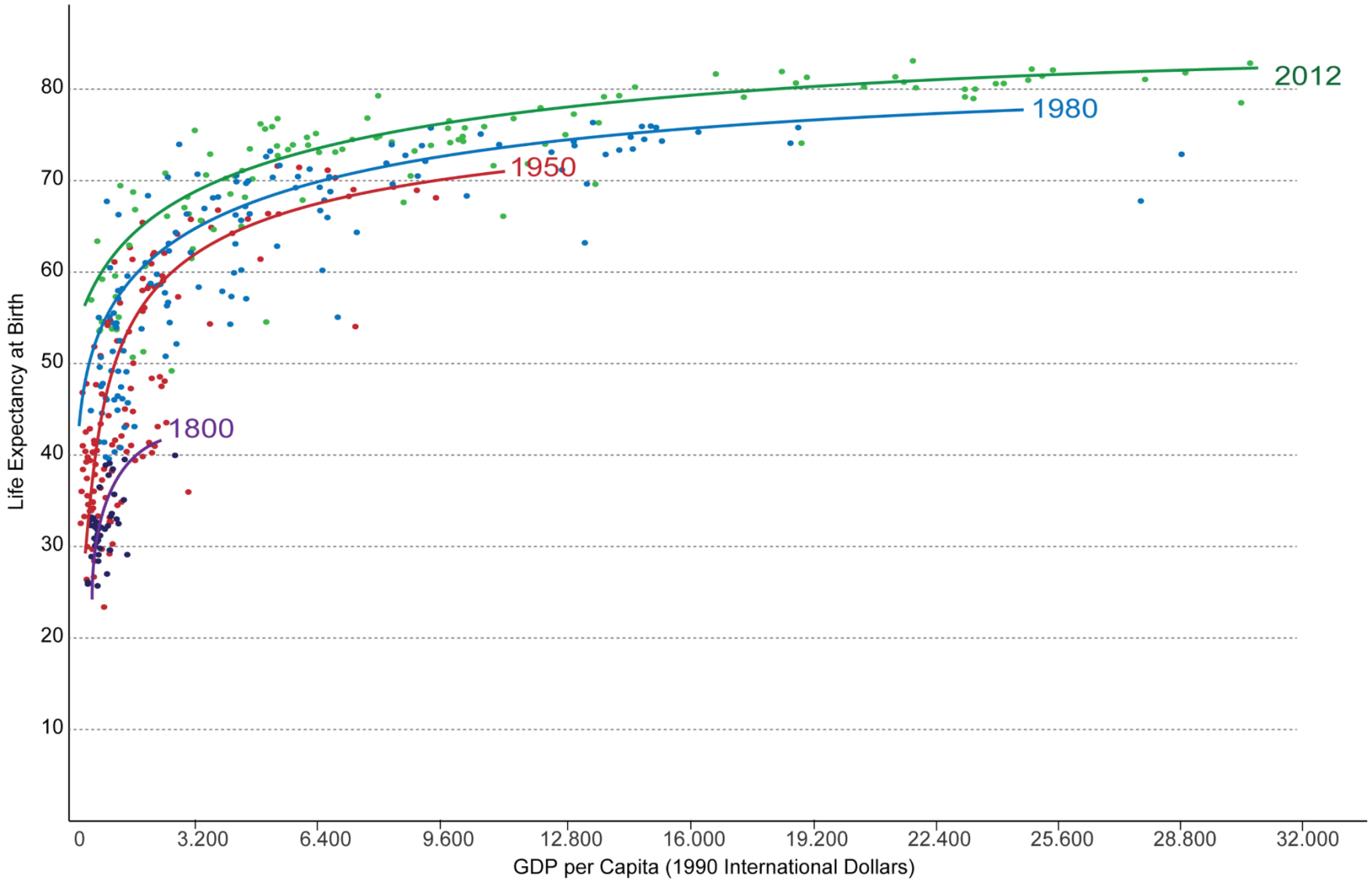


[https://ourworldindata.org/life-expectancy]

Our World in Data

Life Expectancy vs. GDP per Capita from 1800 to 2012 – by Max Roser

GDP per capita is measured in International Dollars. This is a currency that would buy a comparable amount of goods and services a U.S. dollar would buy in the United States in 1990. Therefore incomes are comparable across countries and across time.



This graph displays the correlation between life expectancy and GDP per capita.

Countries with higher GDP have a higher life expectancy, in general.

The relationship seems to follow a logarithmic trend: the unit increase in life expectancy per unit increase in GDP decreases as GDP per capita increases.

Data sources: Data on life expectancy are from Gapminder.org; data on GDP per capita are from the 'New Maddison Project Database'.
The interactive data visualisation is available at OurWorldinData.org. There you find the raw data and more visualisations on this topic.

data-action-lab.com

DATA EXPLORATION