Data Mining & Exploration

(Data Analysis & Visualisation)

Introductie

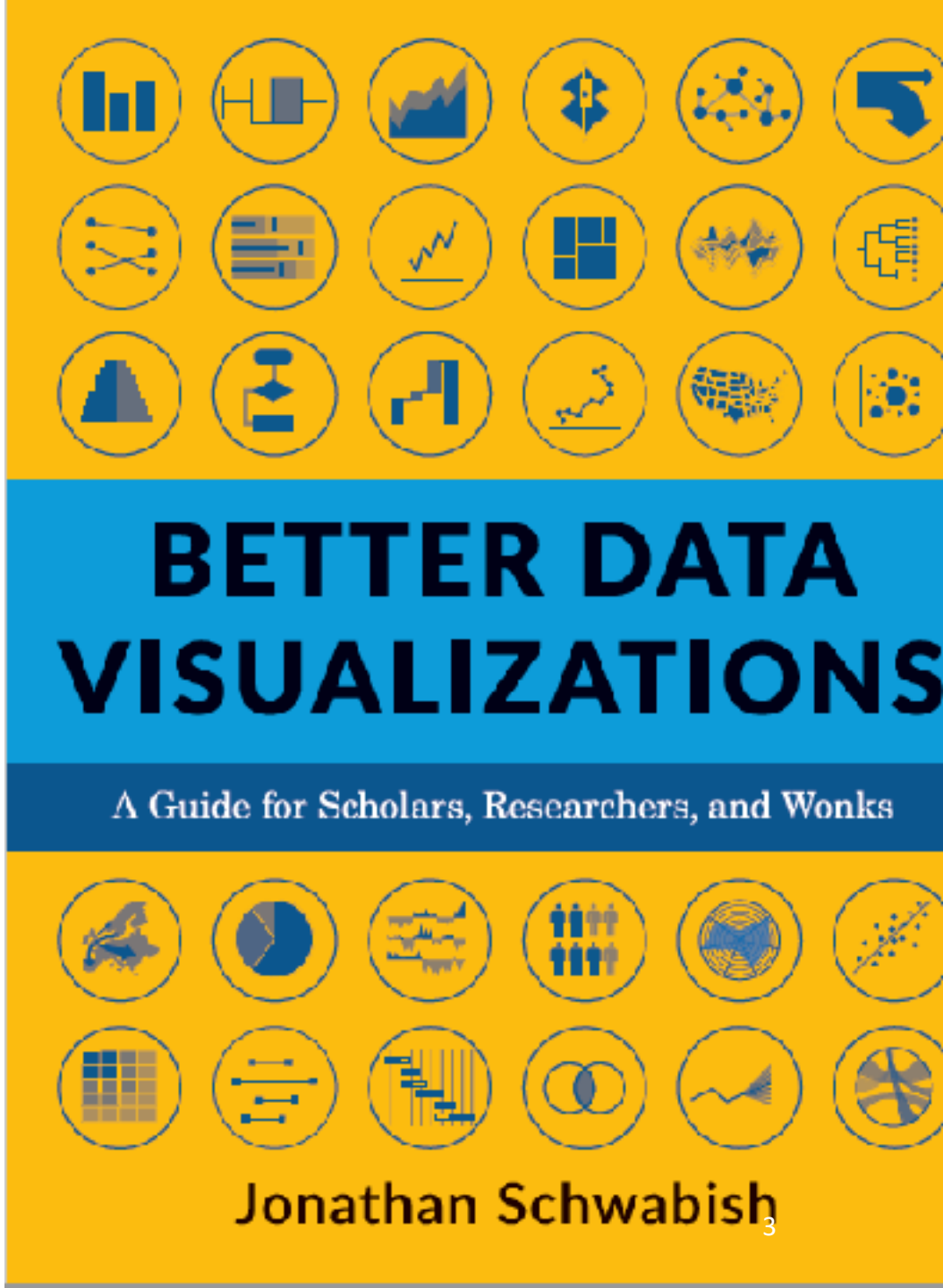
- Raoul Grouls
  raoul.grouls@han.nl
  raoulg

- Kevin Tjoe Ny
  kevin.tjoenij@bd-orange.com

Data Mining & Exploration

- Data analyse
- Reproduceerbaar
  - ETL, analyse, visuals
- Python
  - Git, virtual environments, .py
- Visualisaties

- *“Zelfstandig analyse uitvoeren en een relevant en geloofwaardig verhaal met data en visualisatie vertellen en documenteren.”*
- Better Data Visualizations



Les 1

Better Data Visualizations:

- Gestalt principles
- Preattentive processing
- Five guidelines

Python:

- Setup (local/vm)
- Virtual environment
- Notebook vs .py
- PDM
- Click

Diner (18:00)

Leerdoelen

- De student **begrijpt**:

- Gestalt principles
- Preattentive processing
- Five guidelines
- Virtual Environment

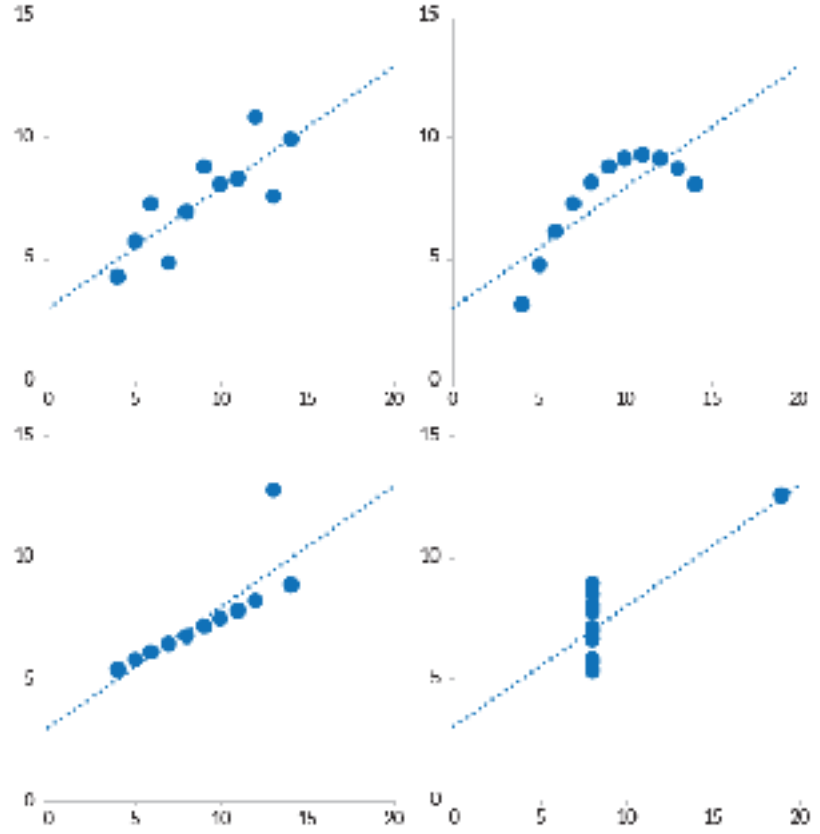
- De student **kan**:

- Gestalt principes & five guidelines toepassen
- Een virtual environment activeren met pdm
- Nieuwe features extraheren met behulp van regular expressions
- Een script vanaf de terminal opstarten
- Click gebruiken voor command line arguments bij een script
- begrijpt de opzet van een project kan een eigen git-repo maken
- Regular expressions toepassen:

Anscombe's Quartet

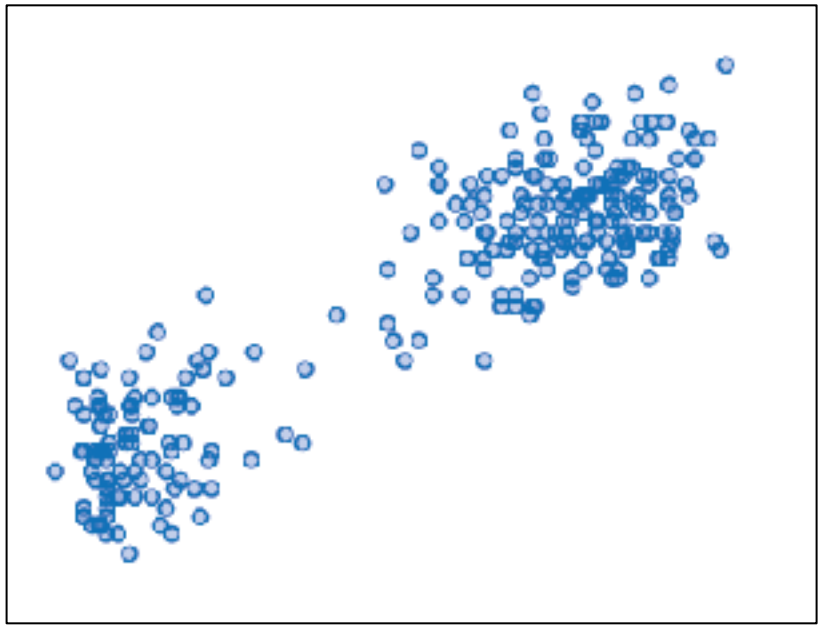
Data set		1	1	2	2	3	3	4	4
Variable		x	y	x	y	x	y	x	y
Obs. No.	1 :	10	8.0	10	9.1	10	7.5	8	6.6
	2 :	8	7.0	8	8.1	8	6.8	8	5.8
	3 :	13	7.6	13	8.7	13	12.7	8	7.7
	4 :	9	8.8	9	8.8	9	7.1	8	8.8
	5 :	11	8.3	11	9.3	11	7.8	8	8.5
	6 :	14	10.0	14	8.1	14	8.8	8	7.0
	7 :	6	7.2	6	6.1	6	6.1	8	5.3
	8 :	4	4.3	4	3.1	4	5.4	19	12.5
	9 :	12	10.8	12	9.1	12	8.2	8	7.6
	10 :	7	4.8	7	7.3	7	6.4	8	7.9
	11 :	5	5.7	5	4.7	5	5.7	8	6.9
Mean		9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance		11.0	4.1	11.0	4.1	11.0	4.1	11.0	4.1
Correlation		0.816		0.816		0.816		0.817	
Regression line		y = 3 + 0.5x		y = 3 + 0.5x		y = 3 + 0.5x		y = 3 + 0.5x	

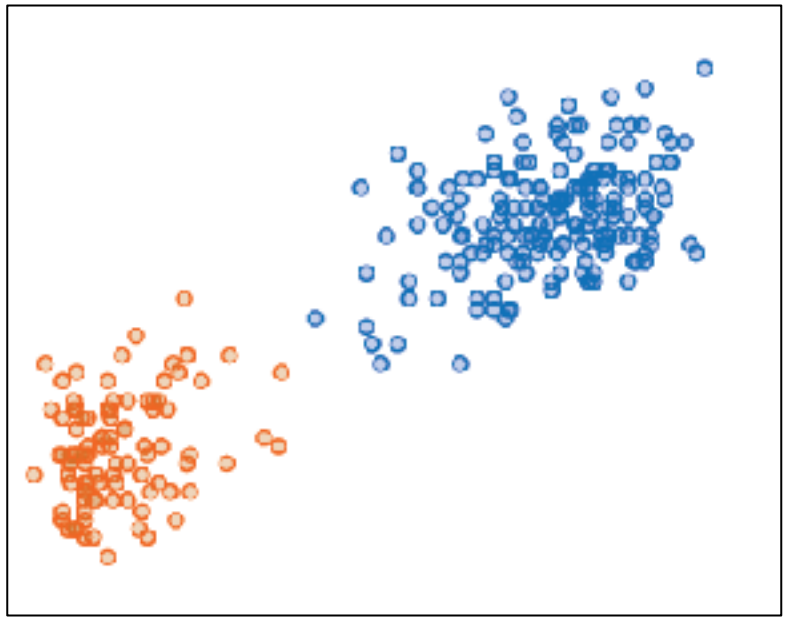
Anscombe’s Quartet

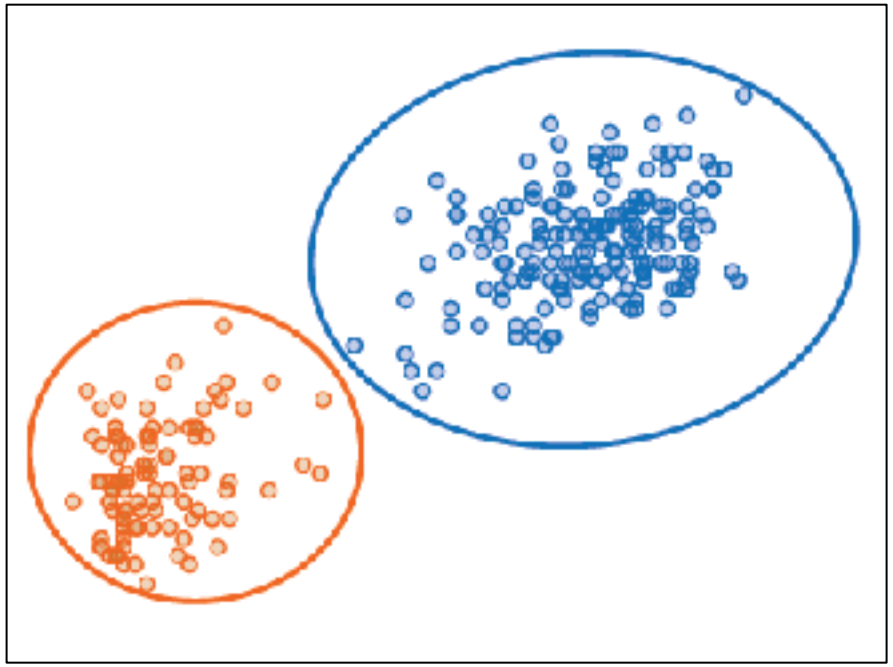


Gestalt Principles (1)

- **Proximity**
- **Similarity**
- **Enclosure**
- Closure
- Continuity
- Connection

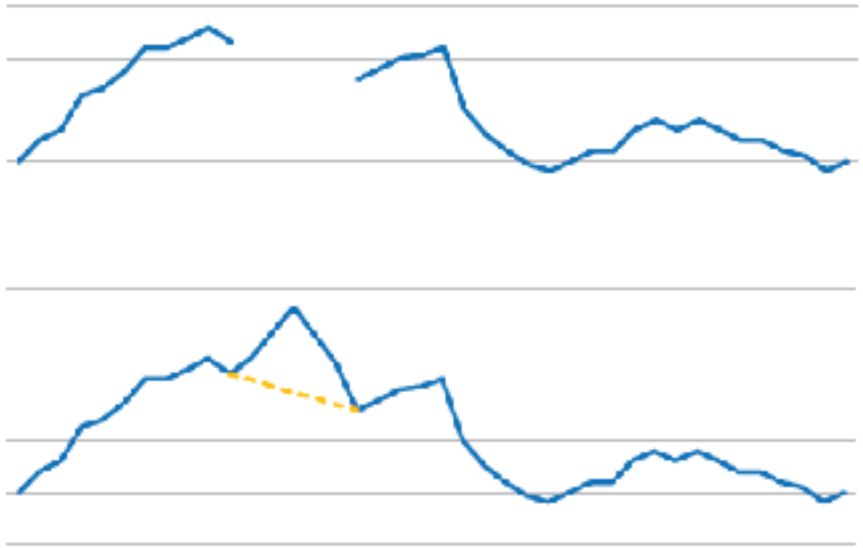


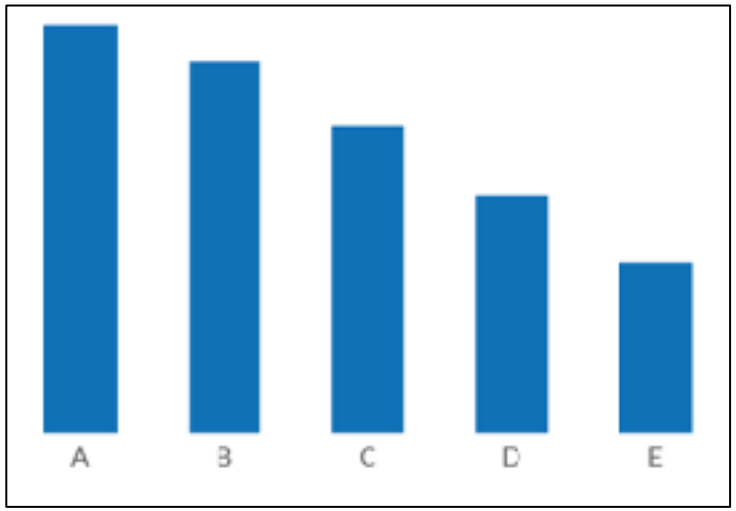


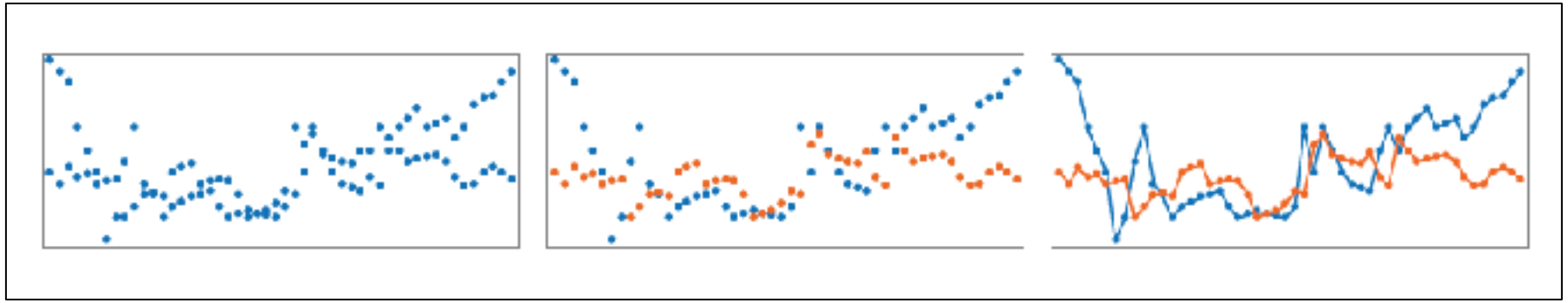


Gestalt Principles (2)

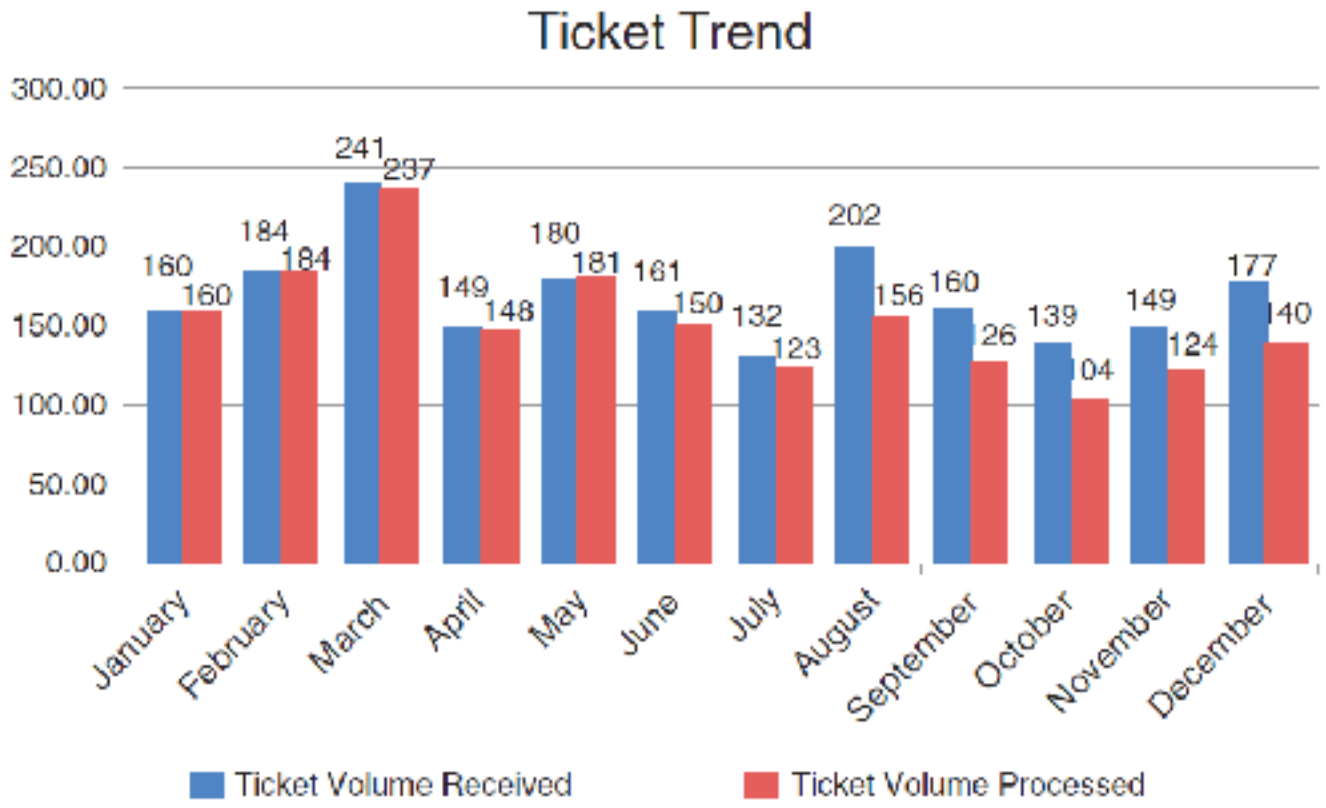
- Proximity
- Similarity
- Enclosure
- **Closure**
- **Continuity**
- **Connection**



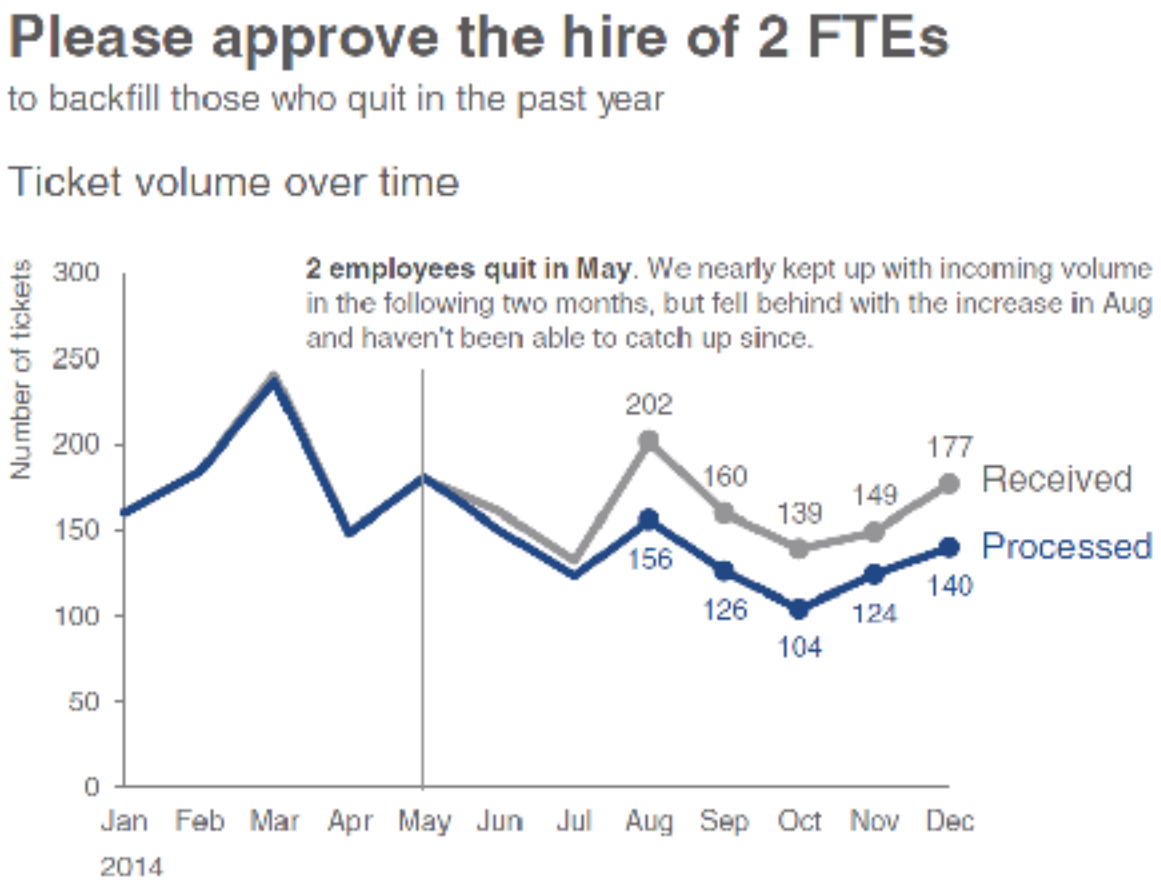


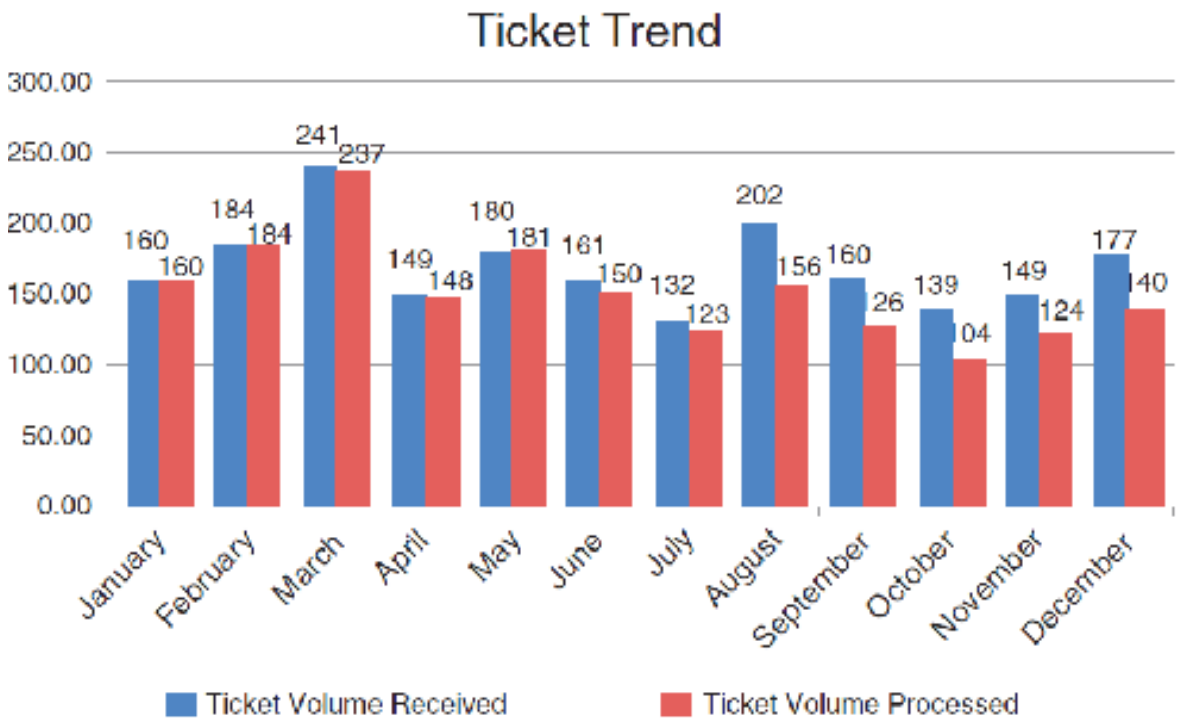


Gestalt Principles



Gestalt Principles

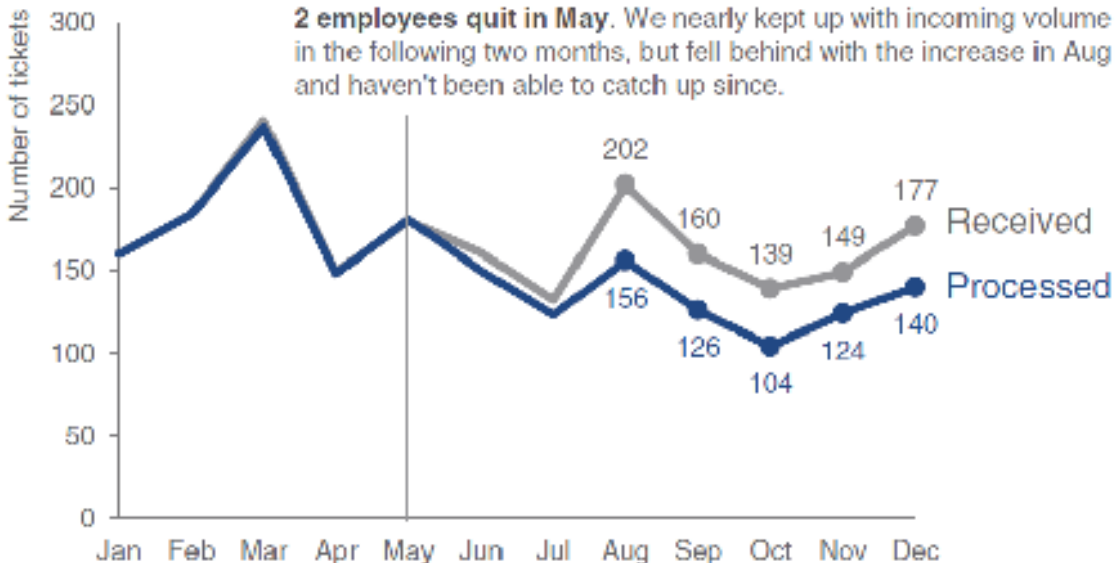




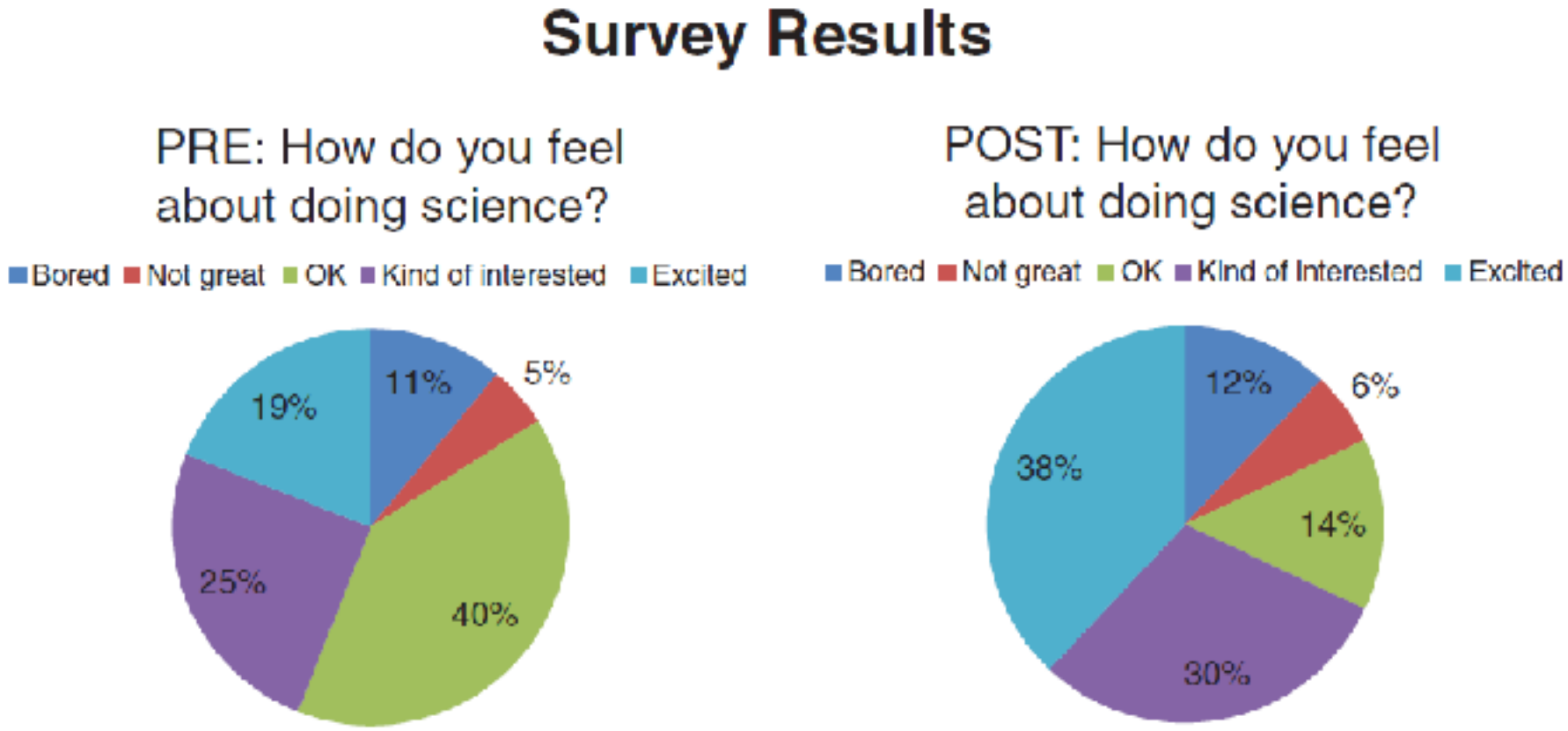
Please approve the hire of 2 FTEs

to backfill those who quit in the past year

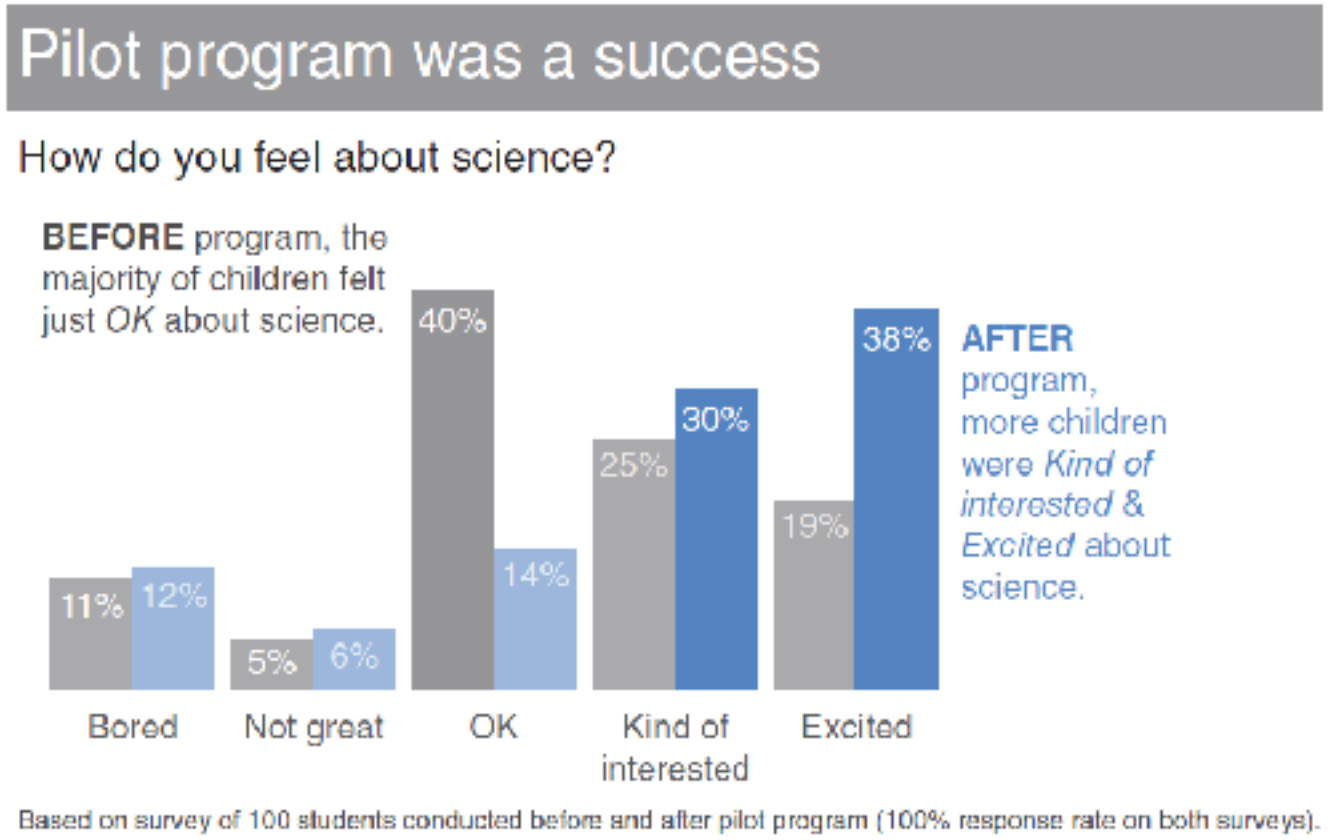
Ticket volume over time



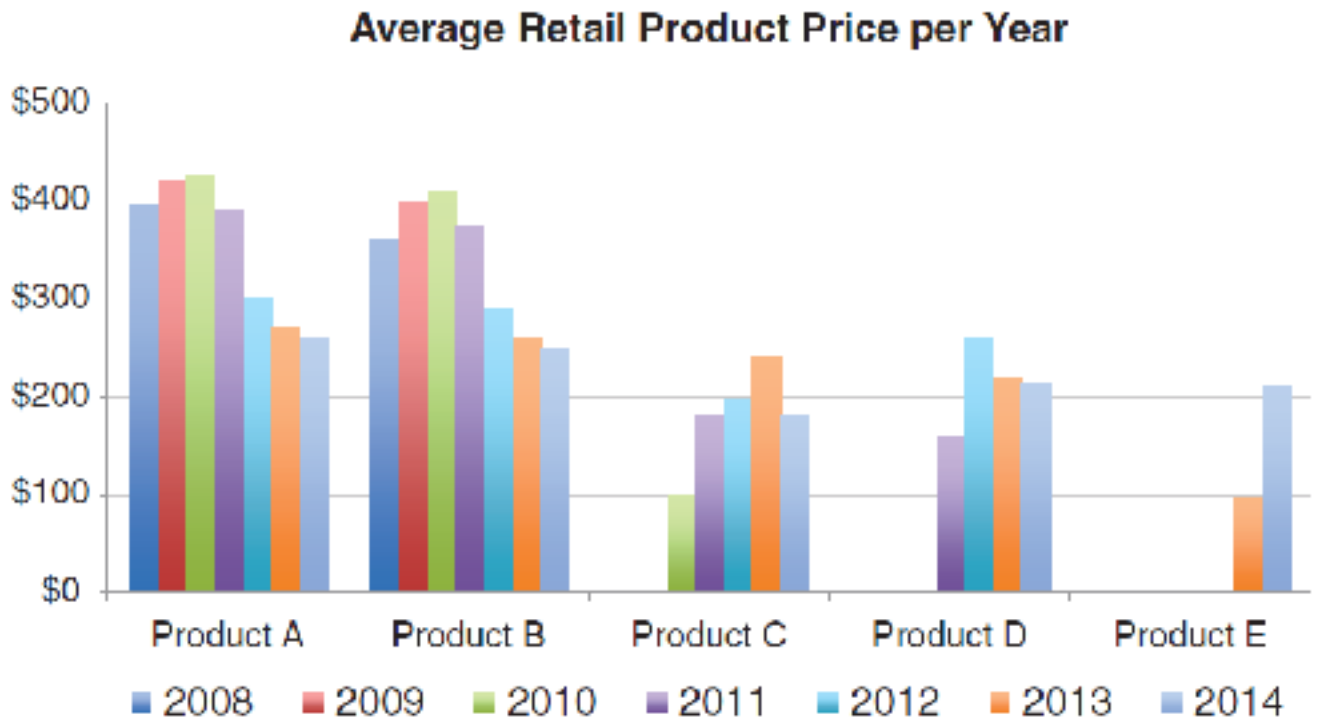
Gestalt Principles



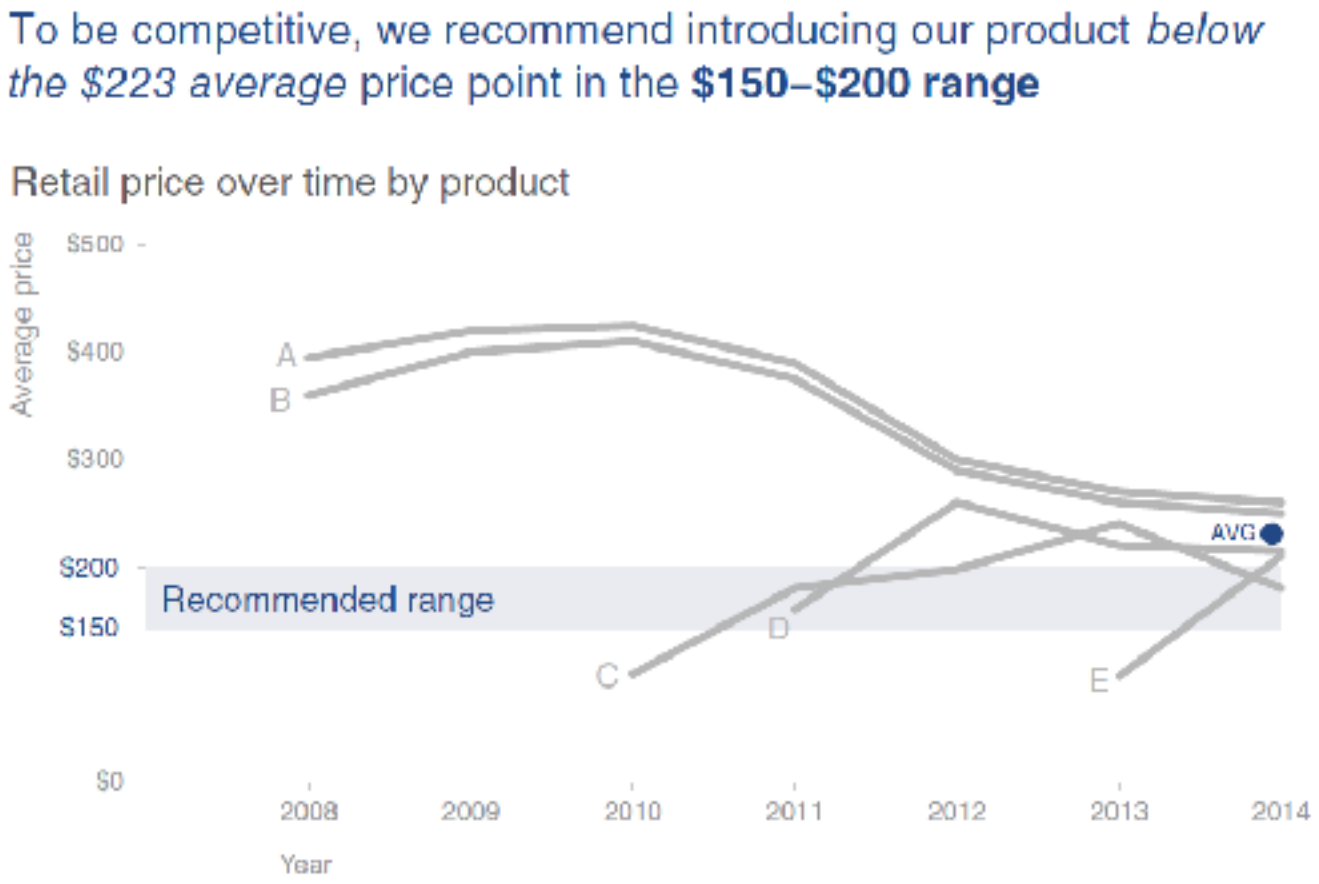
Gestalt Principles



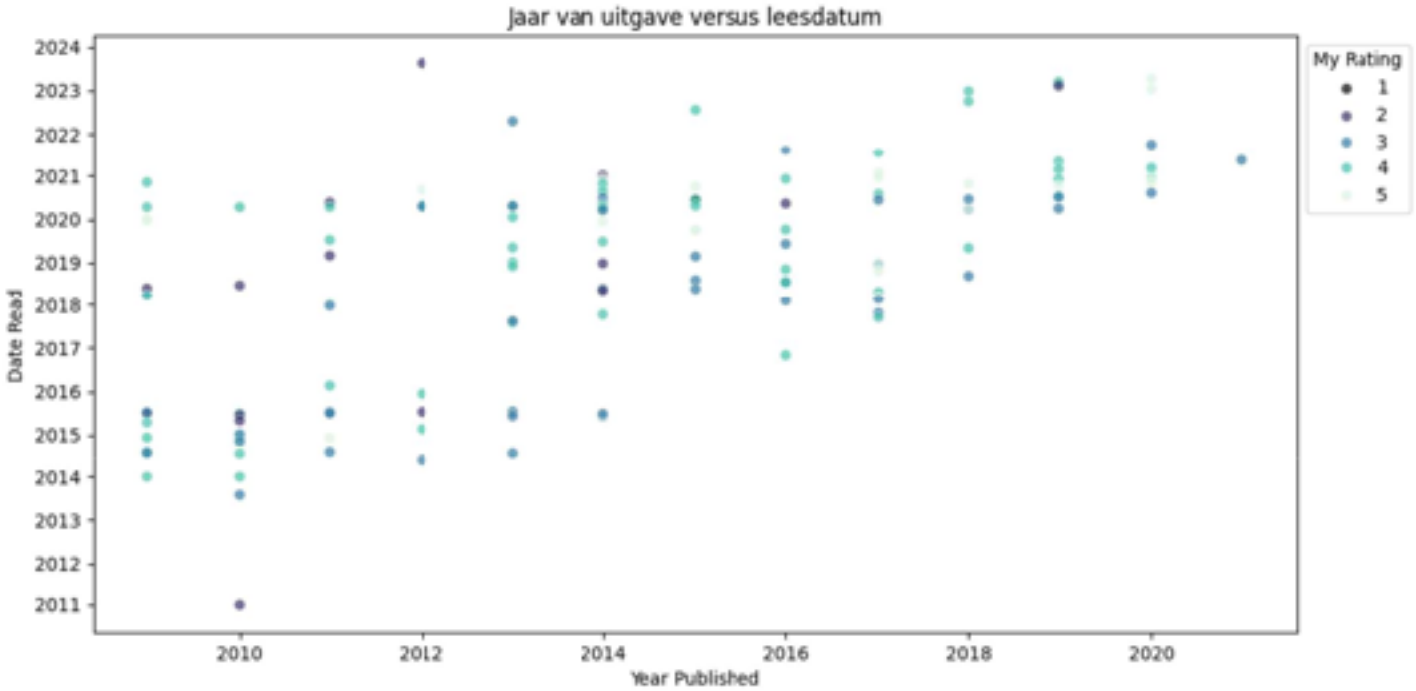
Gestalt Principles



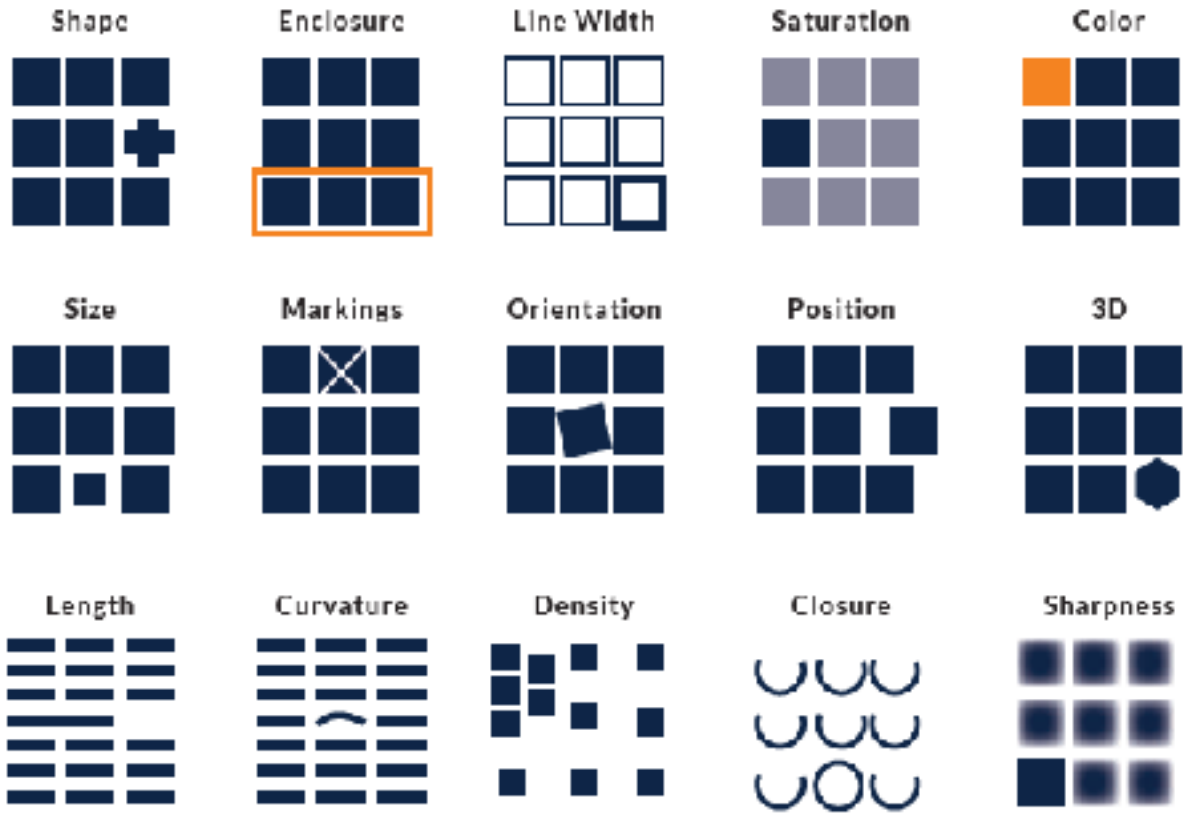
Gestalt Principles



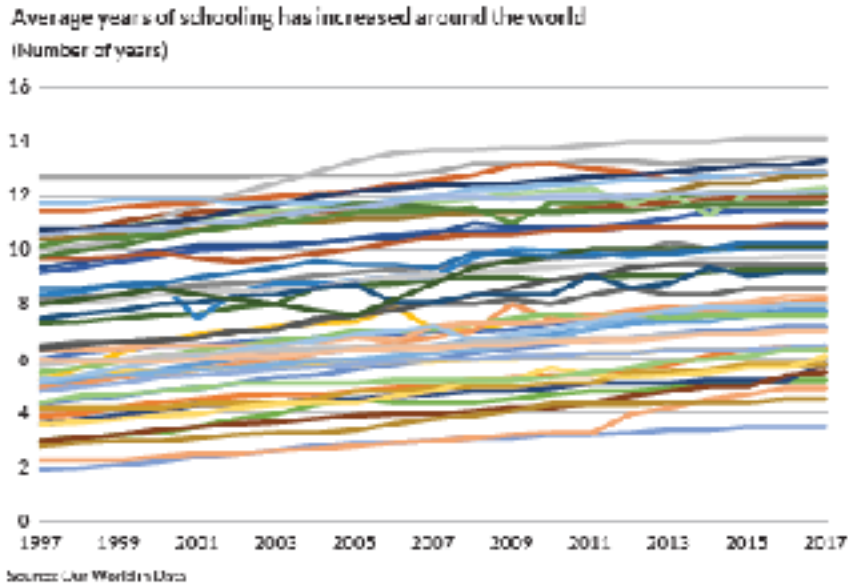
Trend – publicatie jaar vs Date Read

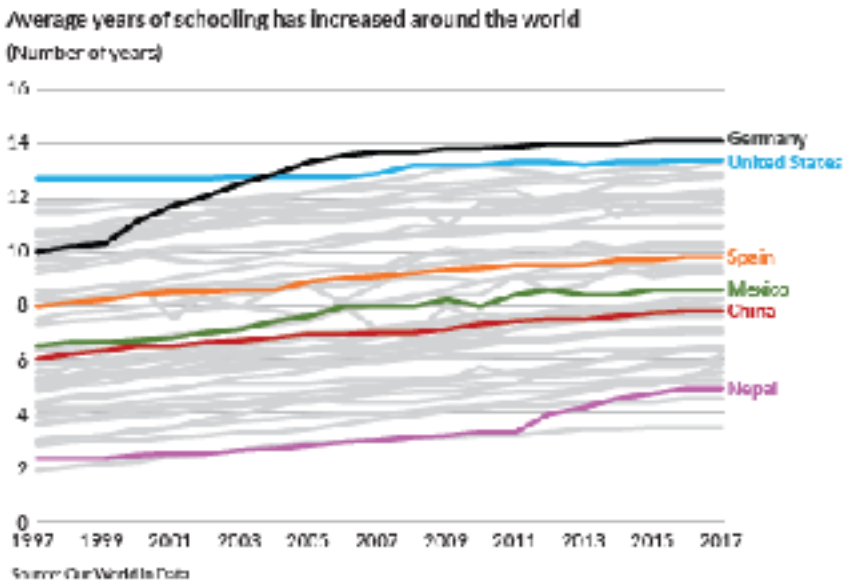


Preattentive Process

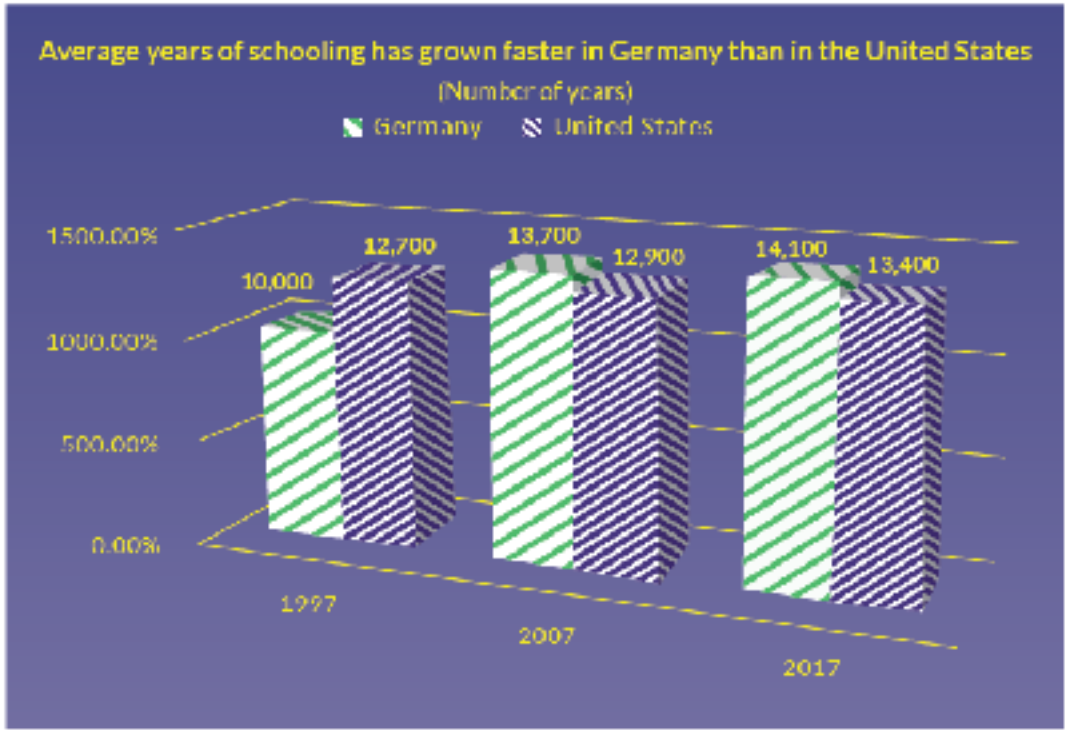


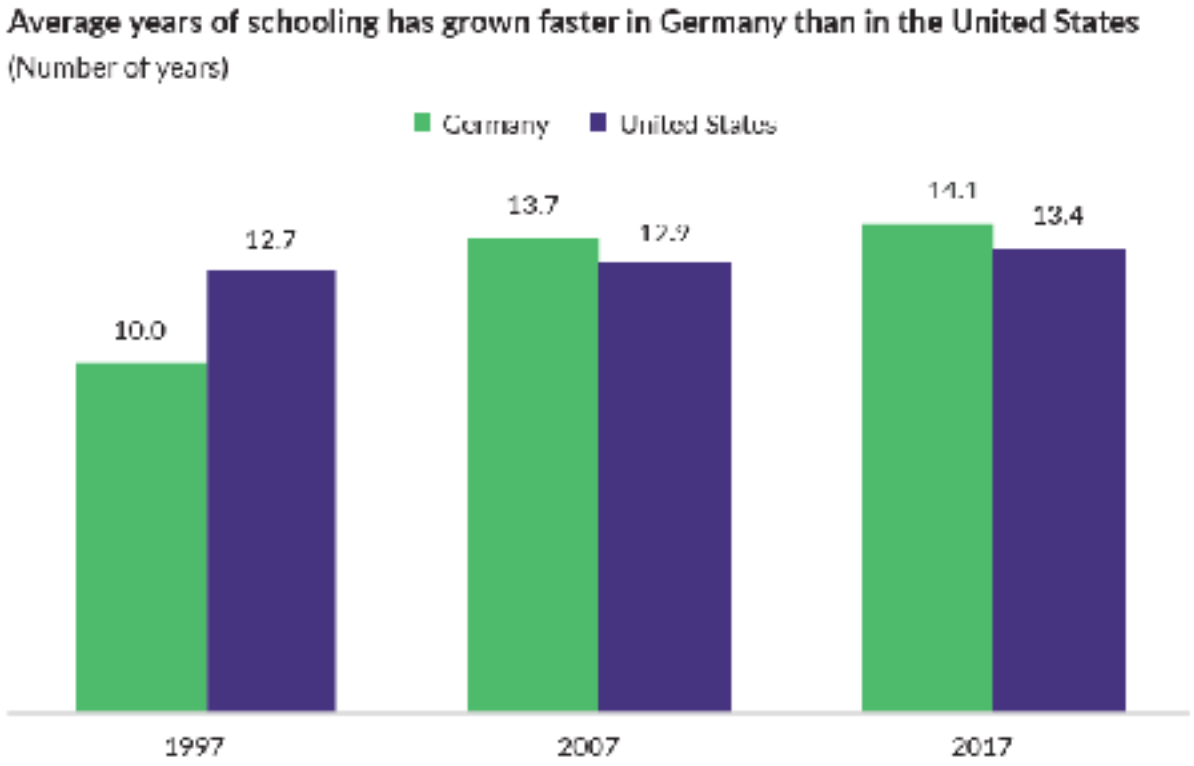
Guidelines – Show the data



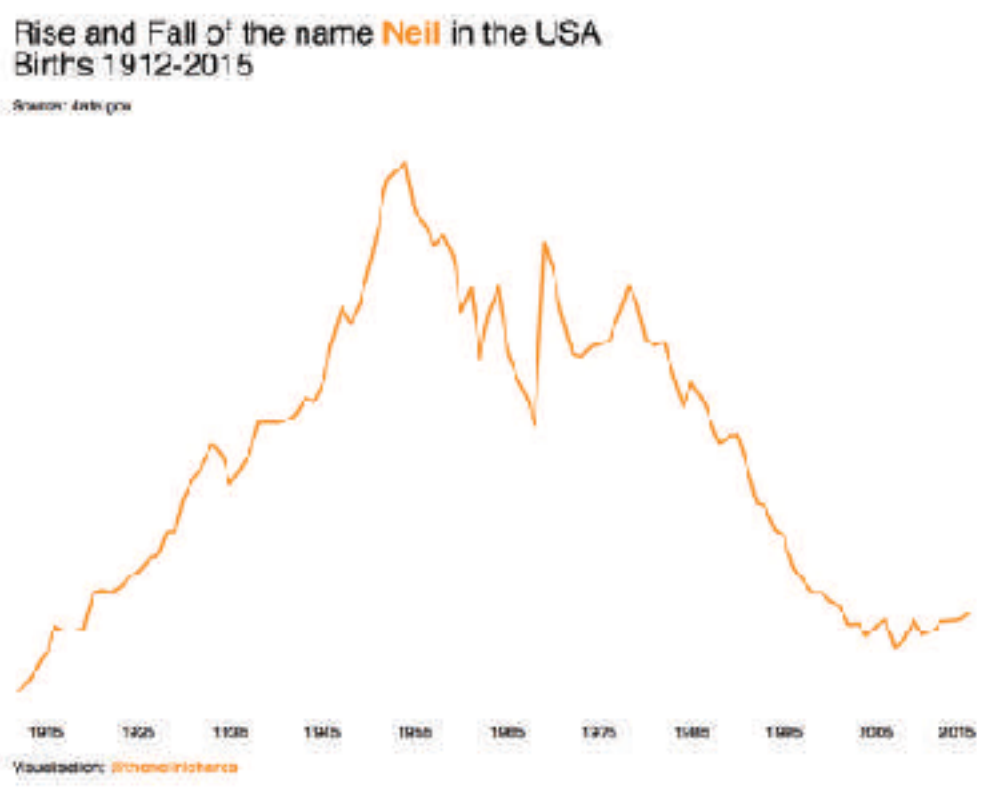


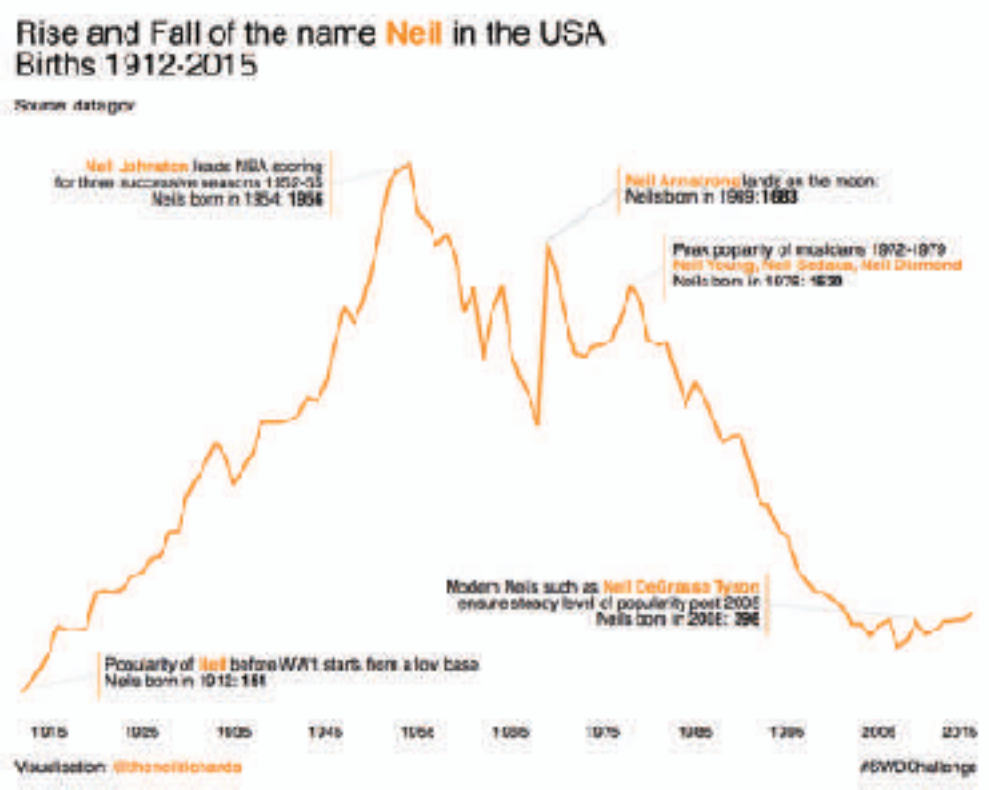
Guidelines – Reduce clutter



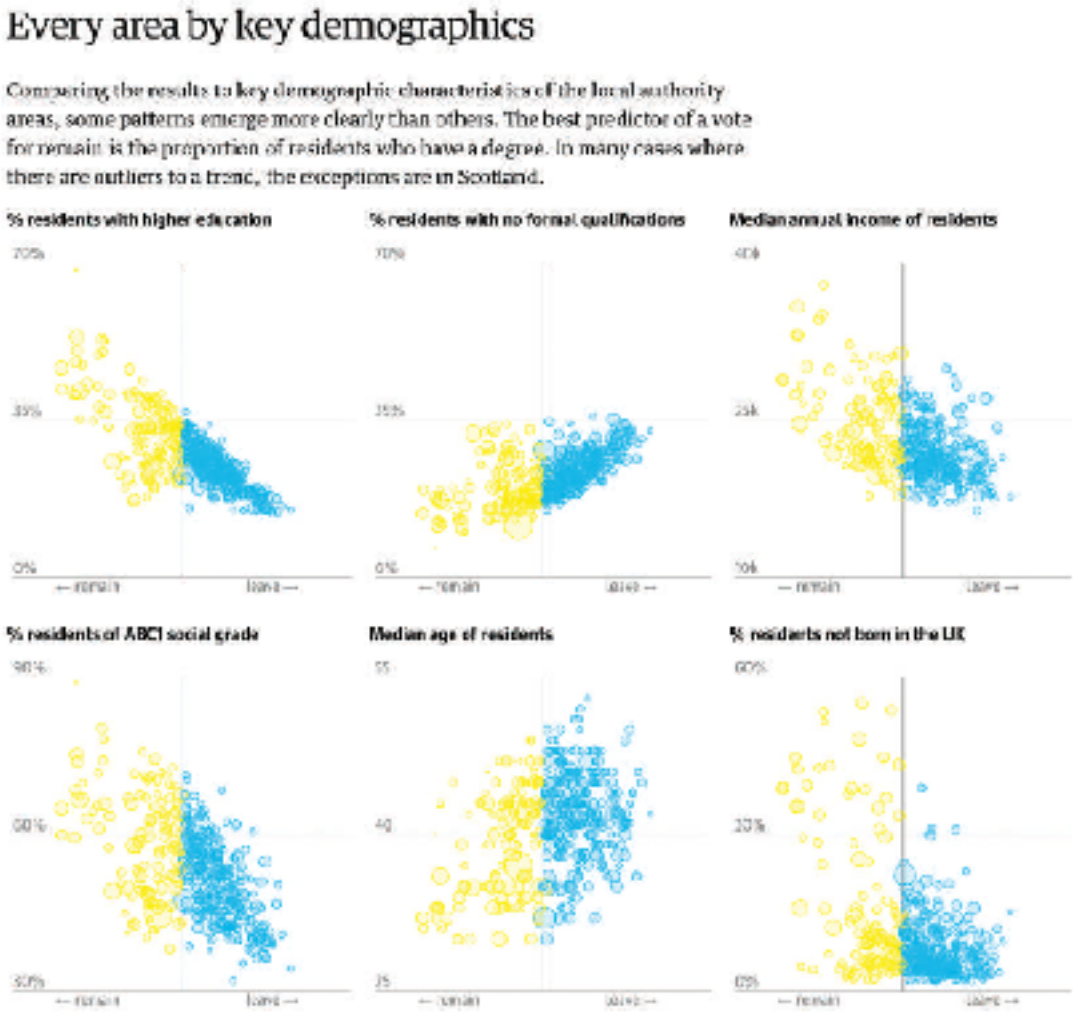


Guidelines – Graphics and text

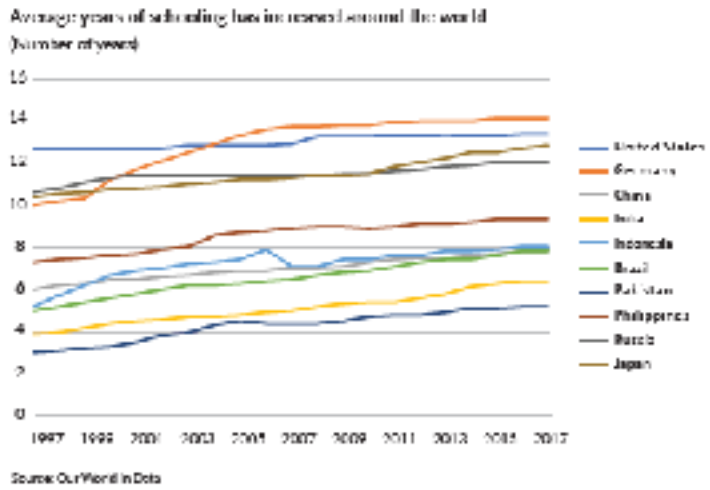


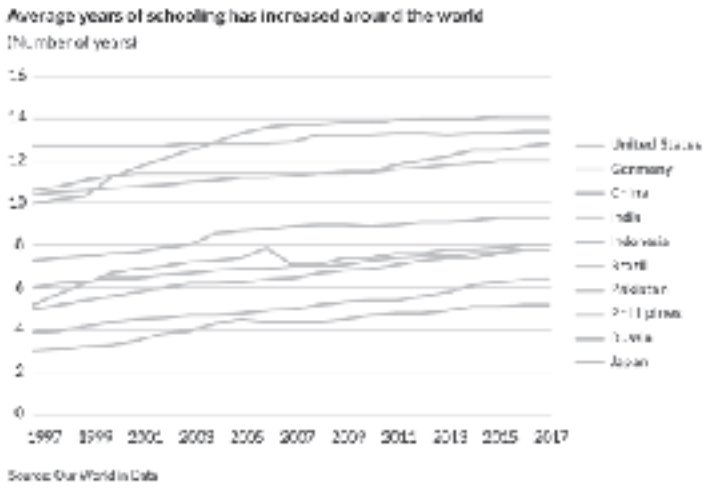


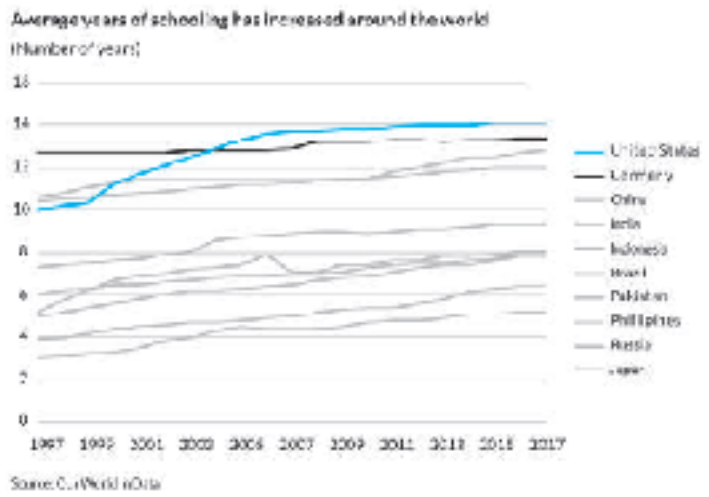
Guidelines – Spaghetti chart



Guidelines – Start with grey







Leerdoelen

- De student **begrijpt**:

- **Gestalt principles**
- **Preattentive processing**
- **Five guidelines**
- Virtual Environment

- De student **kan**:

- **Gestalt principes & five guidelines toepassen**
- Een virtual environment activeren met pdm
- Nieuwe features extraheren met behulp van regular expressions
- Een script vanaf de terminal opstarten
- Click gebruiken voor command line arguments bij een script
- Begrijpt de opzet van een project kan een eigen git-repo maken
- Regular expressions toepassen:

Regex

- Regular Expressions – match character expressions in strings
  - Search for patterns in text, data cleaning and extraction
- import re