Data Analysis and Visualization

Universität Konstanz

Information Visualization

Introduction



Dr. Johannes Fuchs



About me – Lake of Constance



Universität Konstanz

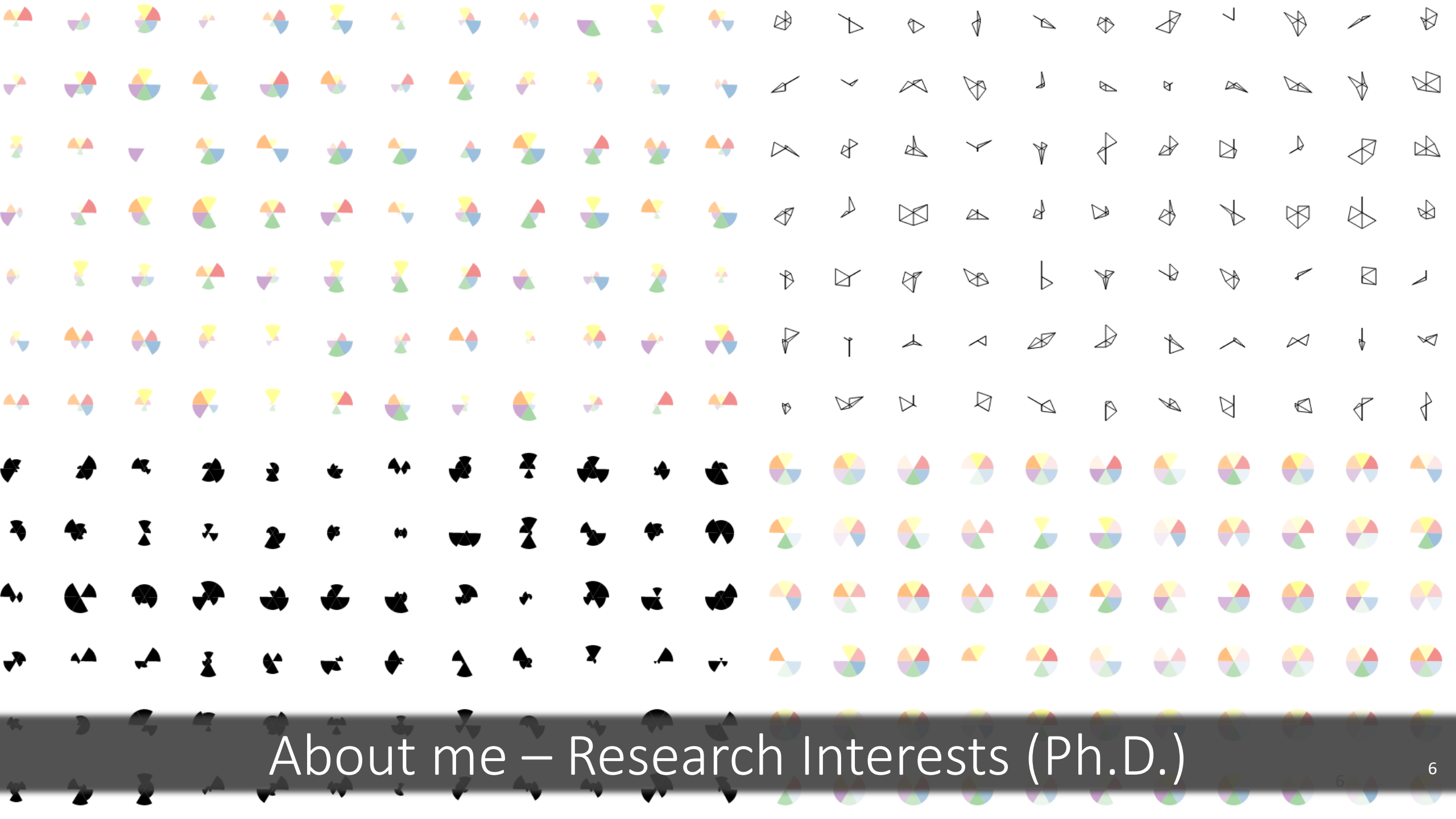
About me – University of Konstanz



About me – Data Analysis and Visualization



About me – Data Analysis and Visualization

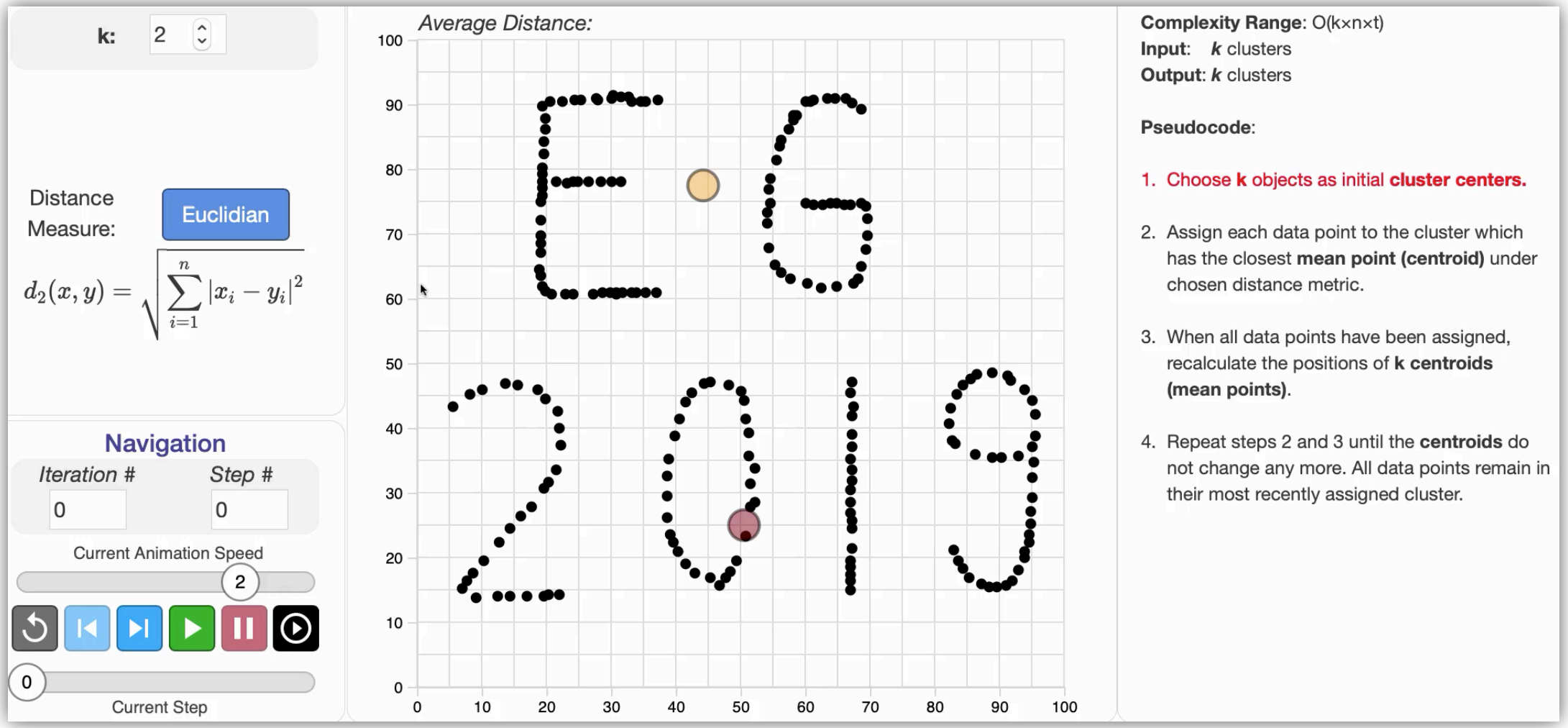


About me – Research Interests (Ph.D.)

Data Analysis
and Visualization



About me – Lecturer since 2015



About me – Research Interests (Education)

https://educlust.dbvis.de/

$Optics$



$DBSCAN$



$k - Means$



About me – Research Interests (Education)

https://educlust.dbvis.de/



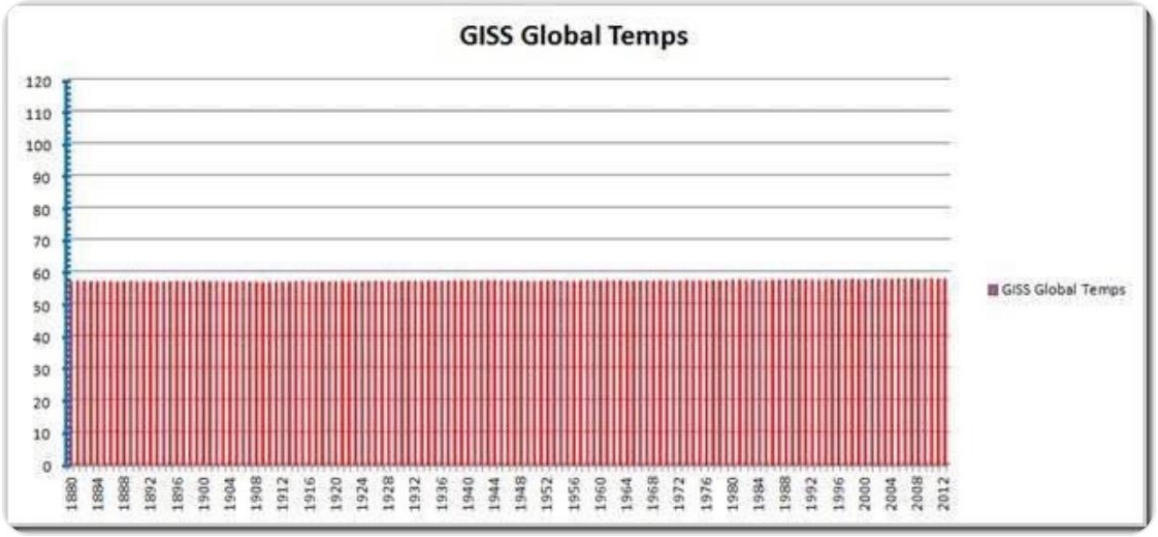
About me – Hobbies



About me – Hobbies

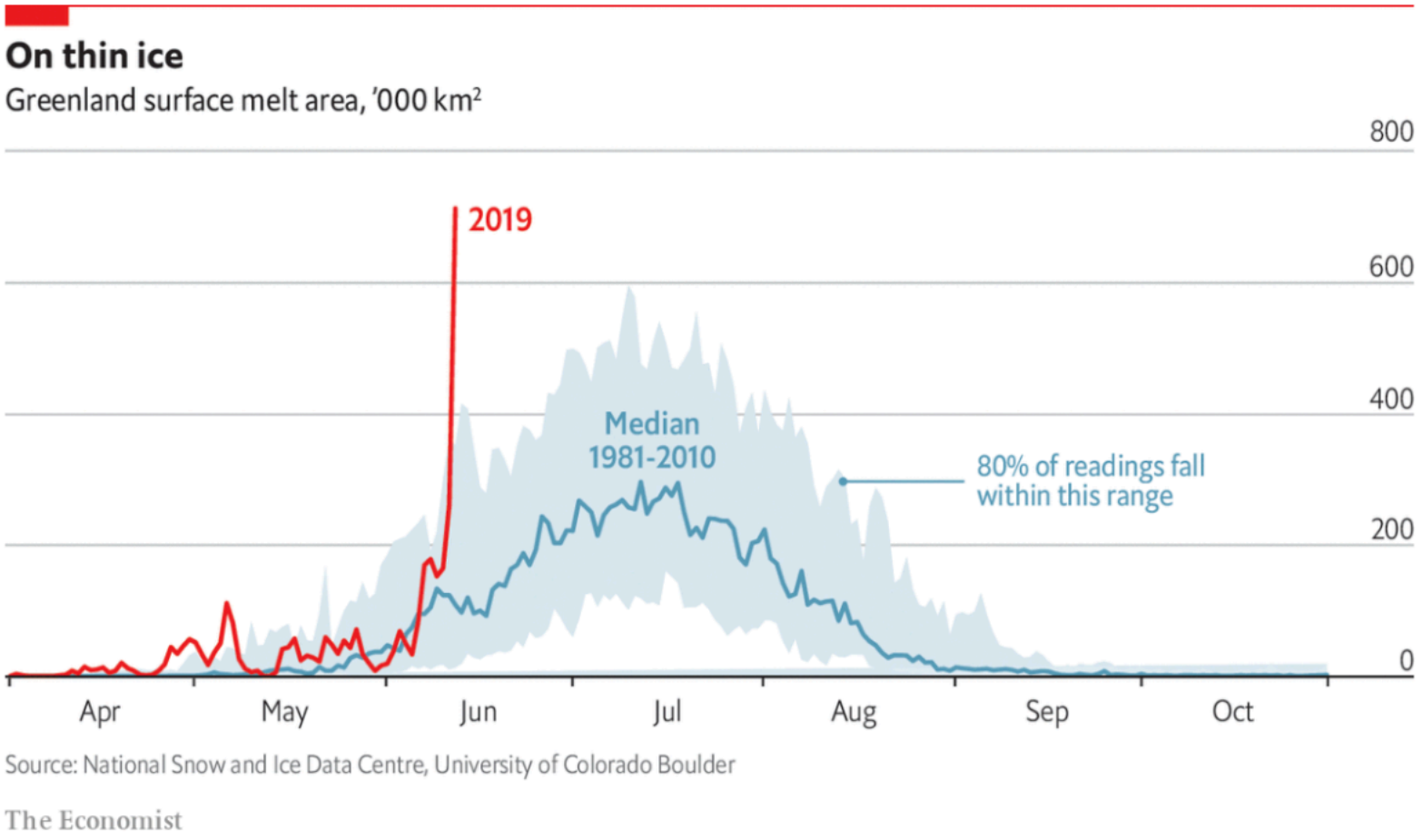
Here's what the "dramatic" rise in global temperature looks like when a realistic scale is used on the Y-axis. It is so not out of the ordinary it is completely ridiculous to make up fake stories about CO2 being the cause of the "climate crisis". The climate is very stable.

Patrick Moore
@EcoSenseNow
https://twitter.com/ecosensenow?lang=de

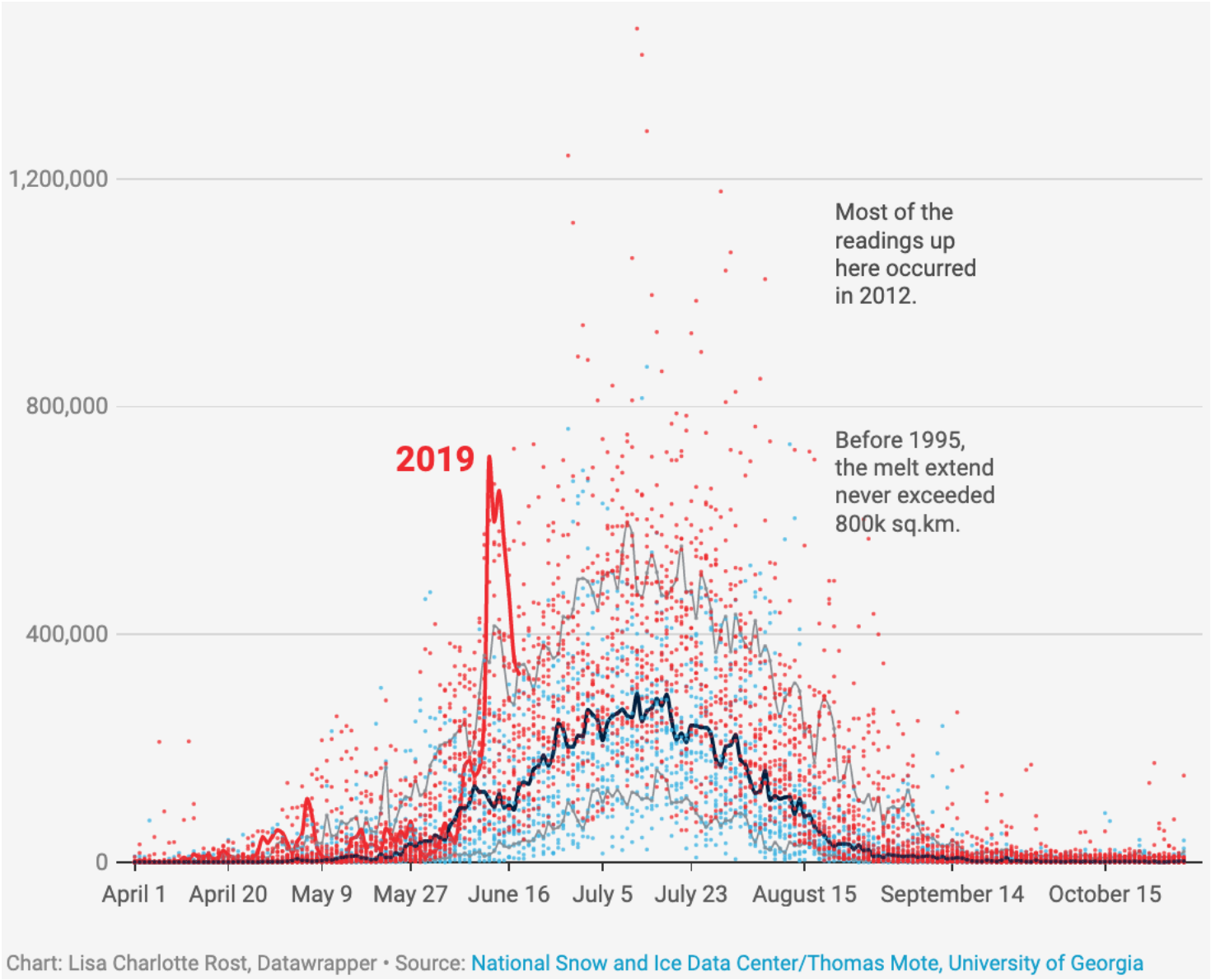


Patrick Moore
@EcoSenseNow
https://twitter.com/ecosensenow?lang=de

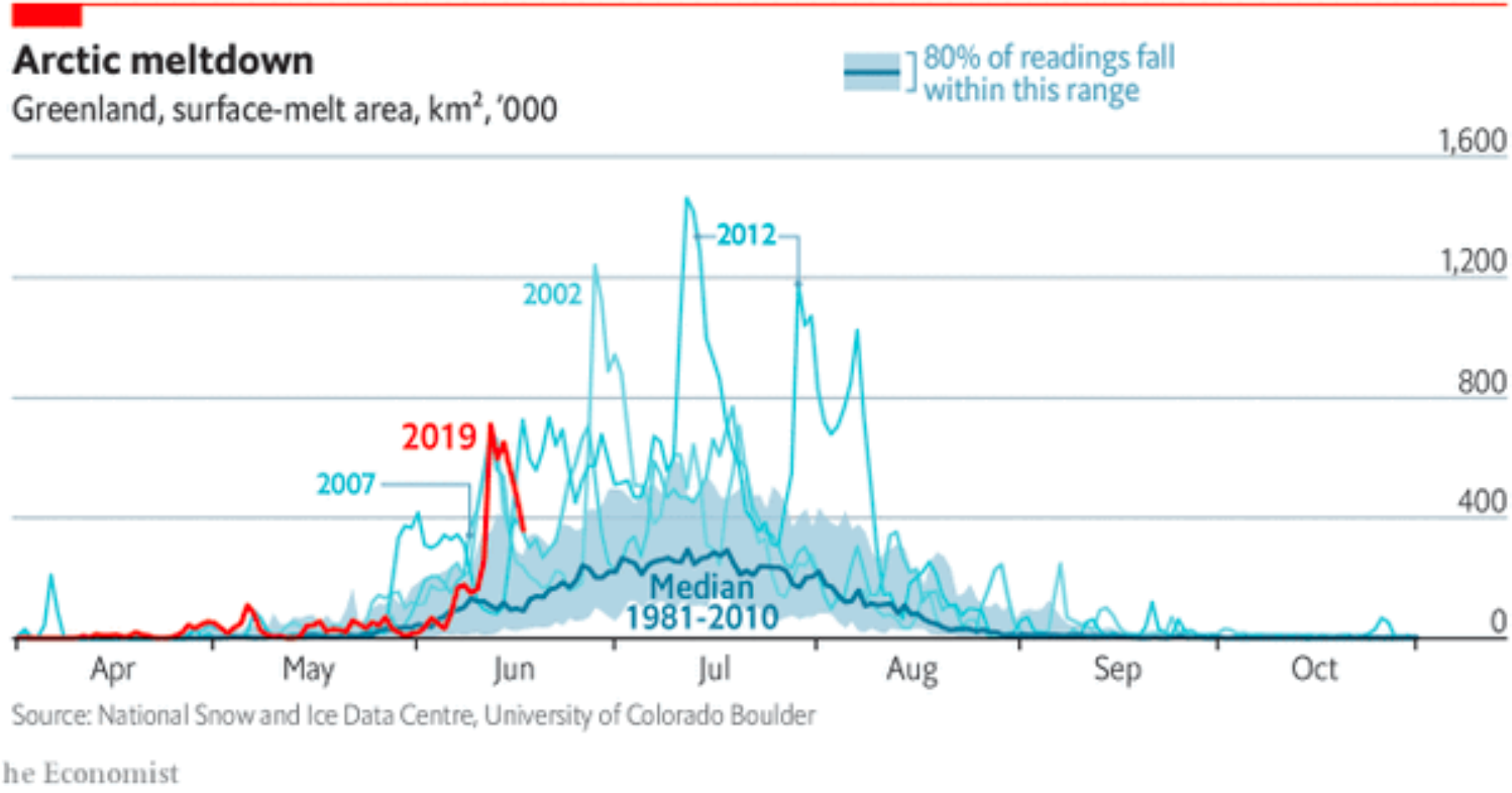
Motivation


On thin ice
Greenland surface melt area, '000 km²
800
600
400
200
0
2019
Median 1981-2010
80% of readings fall within this range
Apr
May
Jun
Jul
Aug
Sep
Oct
Source: National Snow and Ice Data Centre, University of Colorado Boulder
The Economist

Motivation


1,200,000
800,000
400,000
0
Most of the readings up here occurred in 2012.
2019
Before 1995, the melt extend never exceeded 800k sq.km.
April 1
April 20
May 9
May 27
June 16
July 5
July 23
August 15
September 14
October 15
Chart: Lisa Charlotte Rost, Datawrapper • Source: National Snow and Ice Data Center/Thomas Mote, University of Georgia

Motivation


Arctic meltdown
Greenland, surface-melt area, km², '000
80% of readings fall within this range
1,600
1,200
800
400
0
2012
2002
2019
2007
Median
1981-2010
Apr
May
Jun
Jul
Aug
Sep
Oct
Source: National Snow and Ice Data Centre, University of Colorado Boulder
he Economist

Motivation

What do you expect from this seminar?

Introduction

Data Foundation (R)

Coffee break

Visualization Foundation (R)

Scatterplots (R, D3, KNIME)

Lunch break

Scatterplot Matrices & Projections (R)

Coffee break

Parallel Coordinates Plots (R, D3, KNIME)

Conclusions

What we are doing today...



Motivation – Why Visualization?

19

Wife (age)	Husband (age)
45	53
60	55
32	32
33	36
45	48
70	72
50	53
34	38
51	52
40	38
41	43
38	40

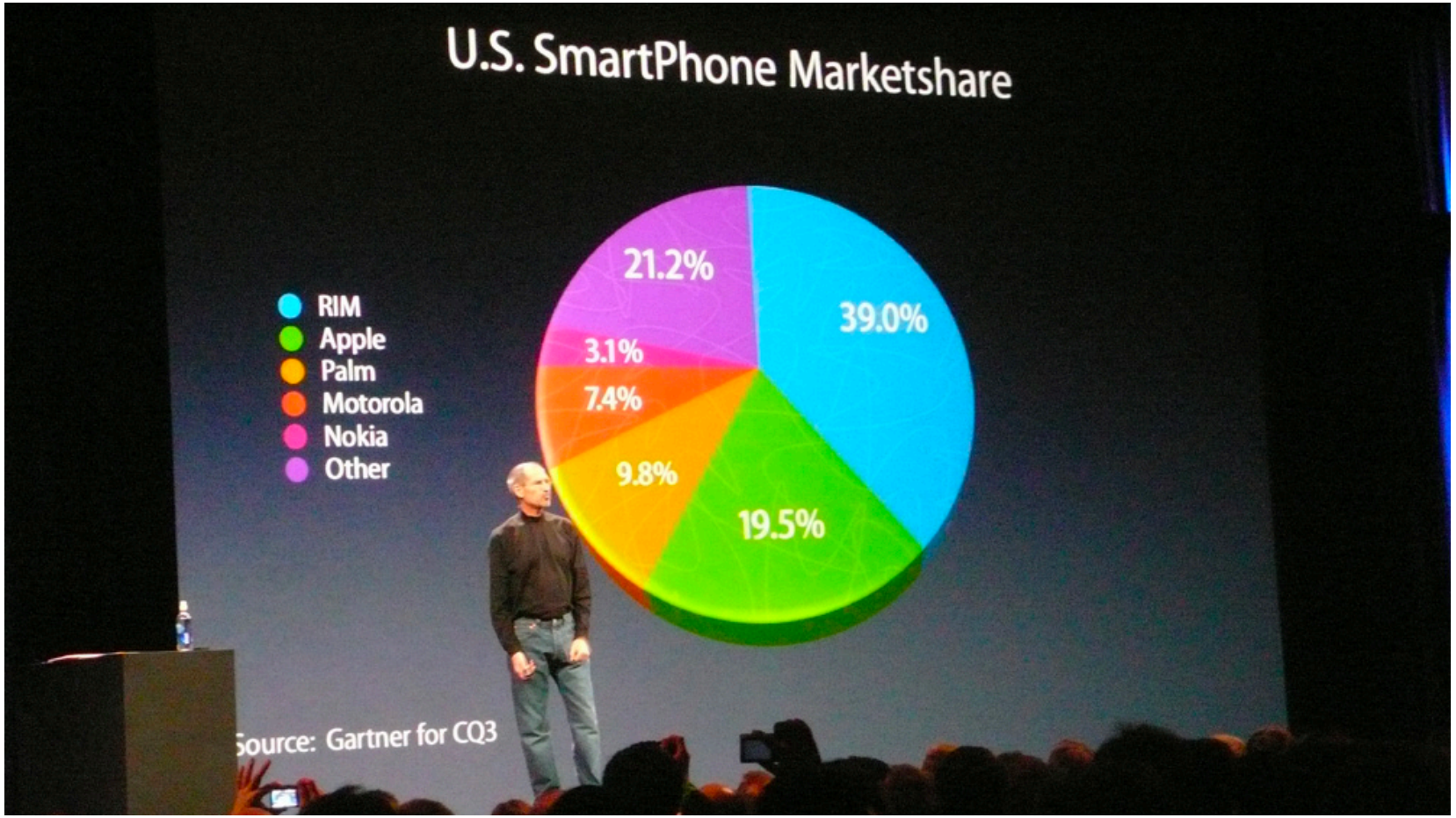
Motivation – Why Visualization?

Wife (age)	Husband (age)
45	53
60	55
32	32
33	36
45	48
70	72
50	53
34	38
51	52
40	38
41	43
38	40



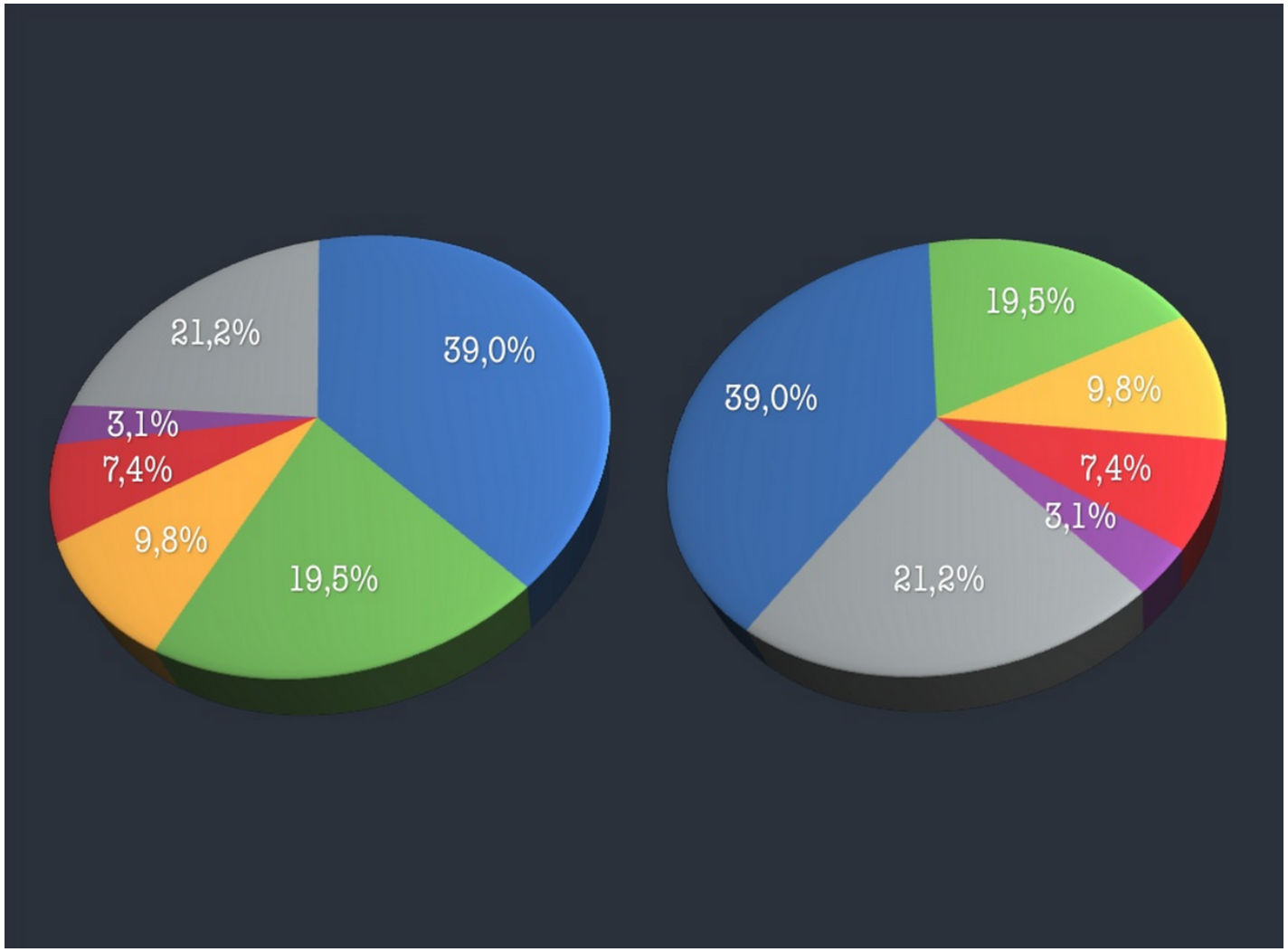
Motivation – Why Visualization?

- Presentation:
  - Starting point: facts to be presented are fixed a priori
  - Process: choice of appropriate presentation techniques
  - Result: high-quality visualization of the data to present facts



Presenting Data

http://stavreas.com/scaling-is-a-race/



Presenting Data

No 3D!



Created by Darkhorse Analytics www.darkhorseanalytics.com

Do not over/under estimate the usefulness of pie charts!

- They are good for simple proportion estimates (relationship of part to the whole).
- They offer at least 4 natural anchor points (0%, 25%, 50%, 75%).
- Space efficient in contrast to a divided bar chart with less anchor points (0%, 50%, 100%).

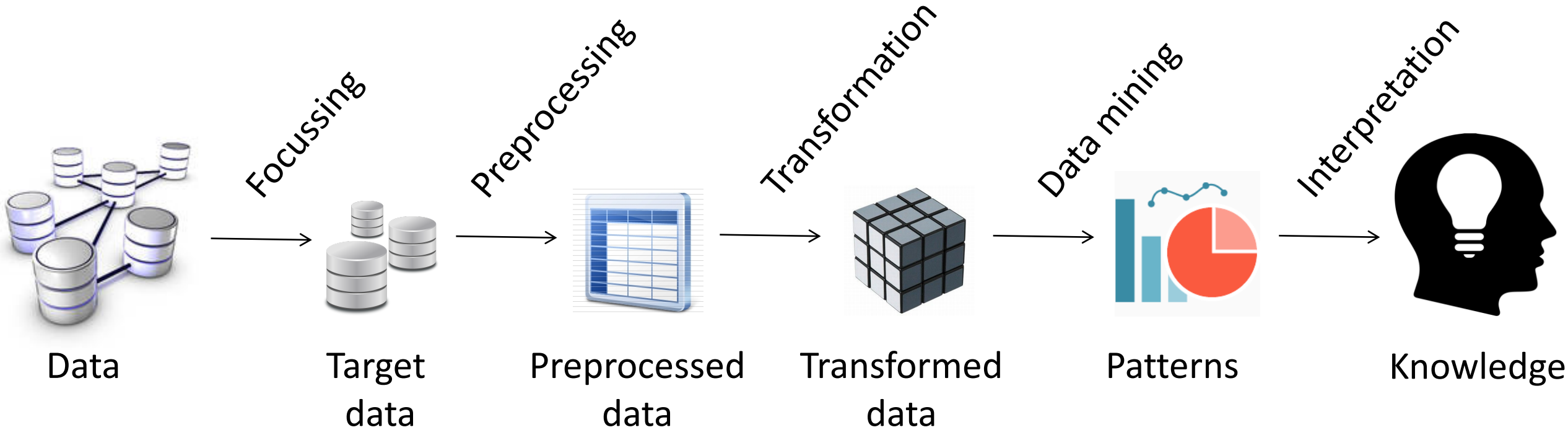
Part-to-whole comparison

Performs worst

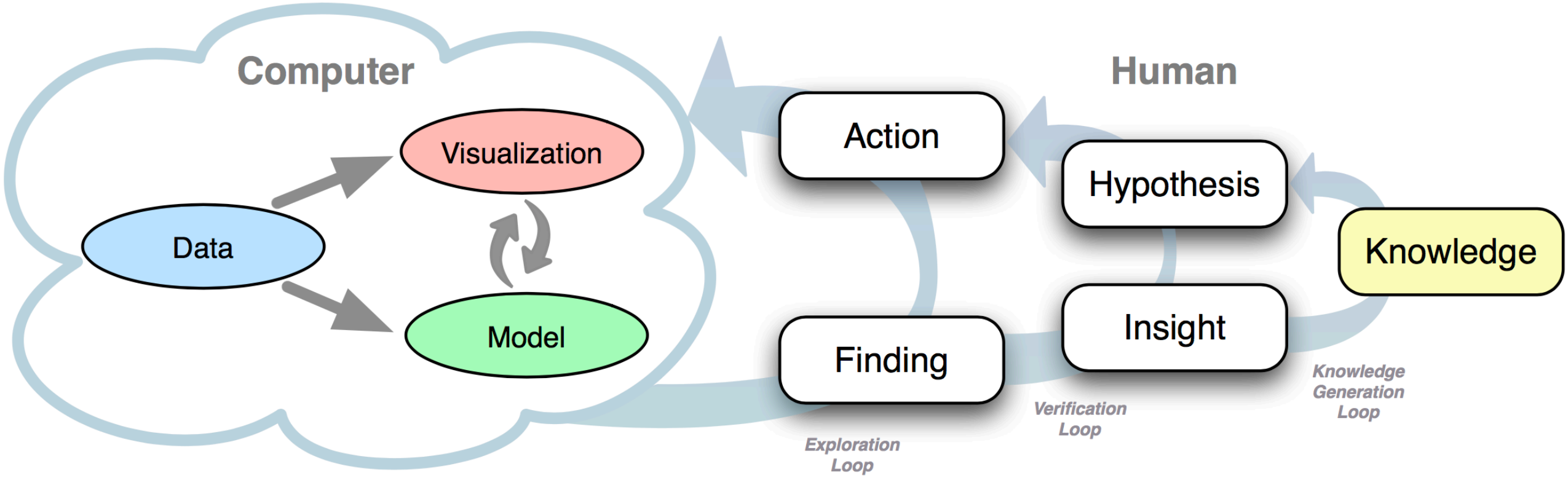


- Center of the pie can be removed without a significant loss in performance

Skau and Kosara, 2016



Motivation – Knowledge Discovery Pipeline



Motivation – Visual Analytics

https://ieeexplore.ieee.org/document/6875967

R-Code: 01_Introduction.r

Set A		Set B		Set C		Set D	
x	**y**	**x**	**y**	**x**	**y**	**x**	**y**
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Hands-On-Session – Anscombe‘s Quartet

Online Demonstration

https://www.autodeskresearch.com/publications/samestats

Motivation – Why Visualization?

- Presentation:
  - Starting point: facts to be presented are fixed a priori
  - Process: choice of appropriate presentation techniques
  - Result: high-quality visualization of the data to present facts
- Confirmatory:
  - Starting point: hypotheses about the data
  - Process: goal-oriented examination of the hypotheses
  - Result: visualization of data to confirm or reject the hypotheses
- Exploratory Analysis:
  - Starting point: no hypotheses about the data
  - Process: interactive, usually undirected search for structures, trends
  - Result: visualization of data to lead to hypotheses about the data

Goals of Visualization

- Visualize data at different steps in the analysis pipeline
- No 3D
- Use bar charts for value comparisons
- Pie charts should only be used for part-to-whole comparisons
- Donut charts and pie charts perform equally well

Guidelines - Introduction