Data visualization as an engineering task

a methodological approach towards creating effective data visualization

Boris Gorelik

http://gorelik.net

AUTOMATTIC

WordPress.com

AZRIELI College of Engineering Jerusalem

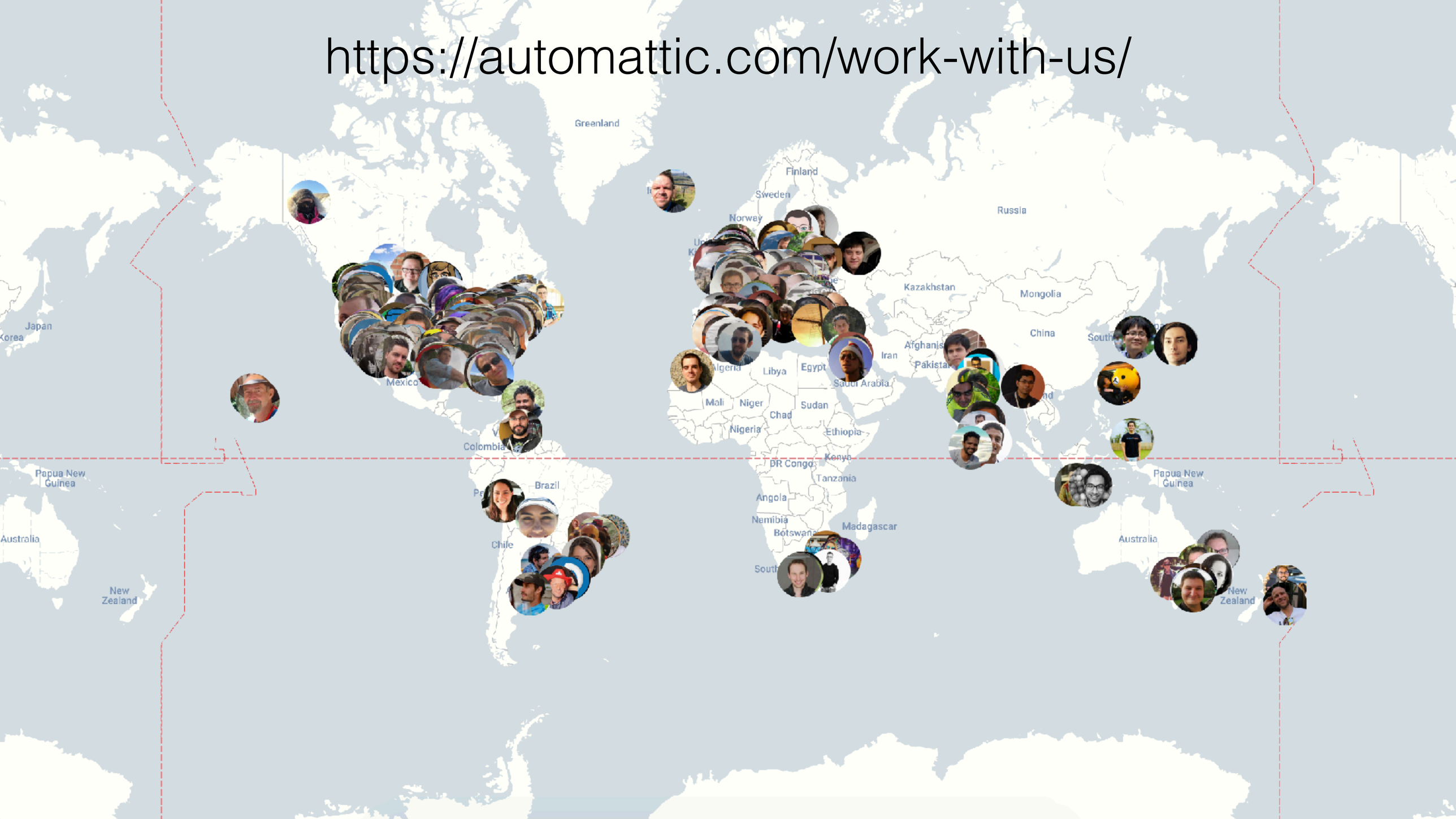
Three most common data visualization mistakes

1

2

3

https://automattic.com/work-with-us/



Three most common data visualization mistakes

1

Why?

Explore

Explain

Explore

Learn

Explain

Teach

Explore

Learn

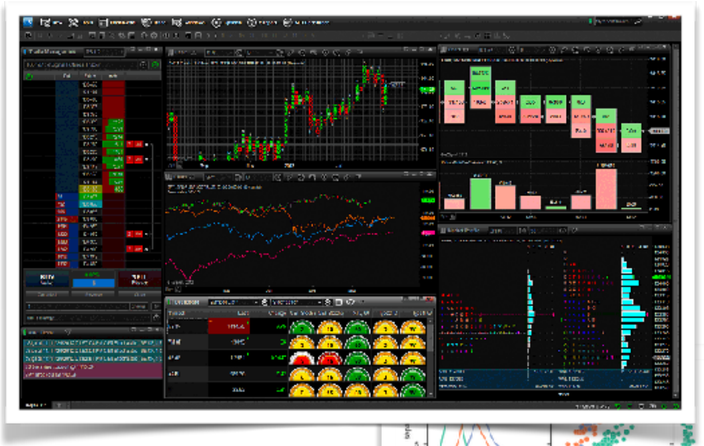
Understand

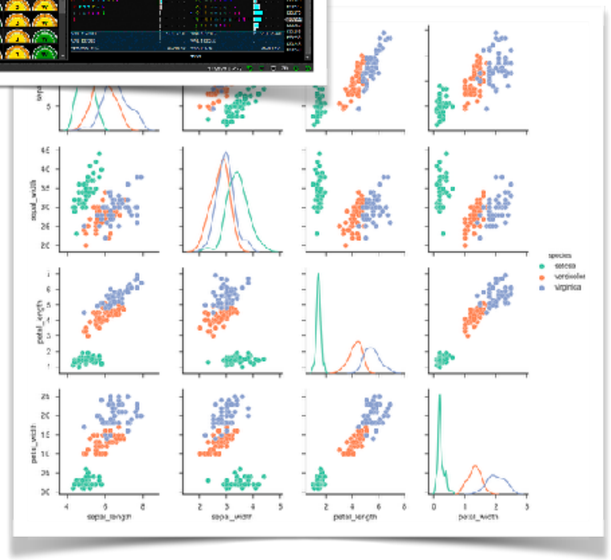
Explain

Teach

Convince

Exploratory

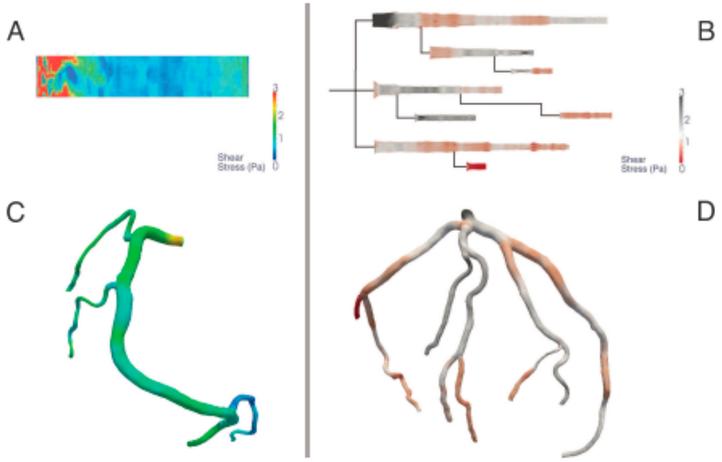




Explanatory

Evaluation of Artery Visualizations for Heart Disease Diagnosis

Michelle A. Borkin, *Student Member, IEEE*, Krzysztof Z. Gajos, Amanda Peters, Dimitrios Mitsouras,
Simone Melchionna, Frank J. Rybicki, Charles L. Feldman, & Hanspeter Pfister, *Senior Member, IEEE*



Evaluation of

ease Diagnosis

Michelle A. Borkin,
Simone Melchionna,

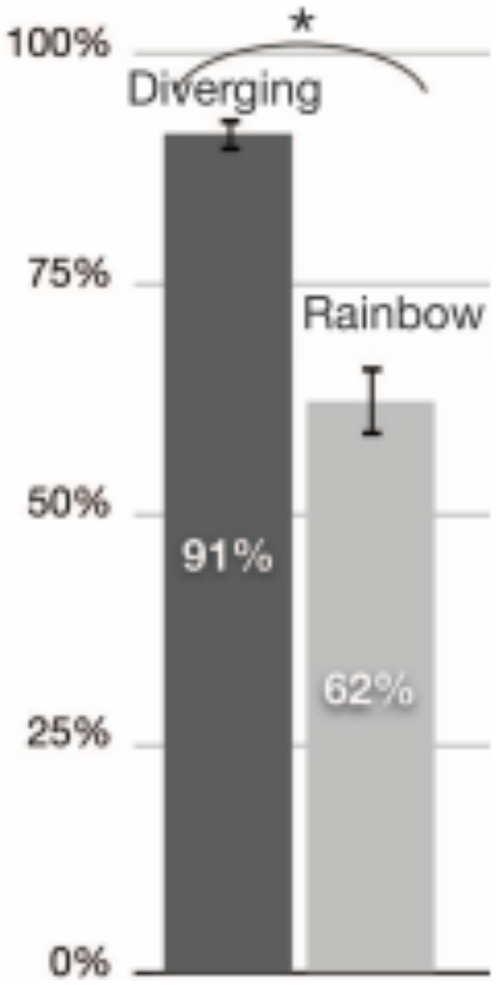
s, Dimitrios Mitsouras,
Senior Member, IEEE

A

B

C

D



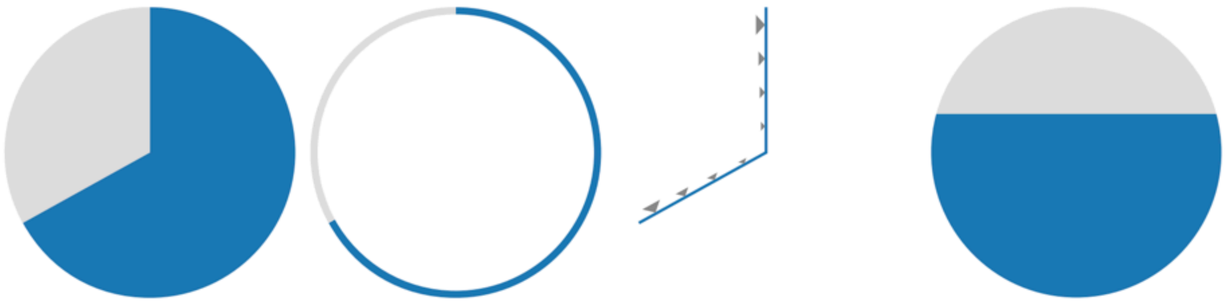
Eurographics Conference on Visualization (EuroVis) 2016
K.-L. Ma, G. Santucci, and J. van Wijk
(Guest Editors)

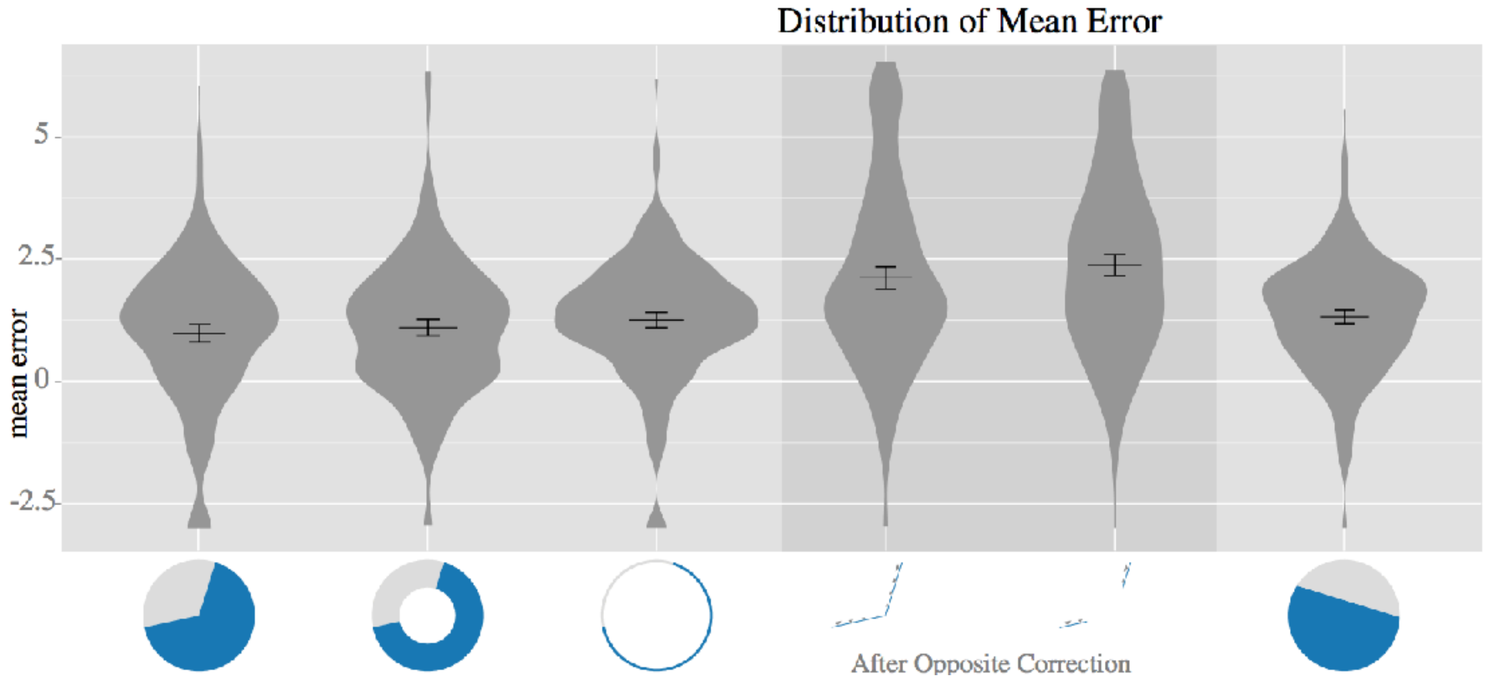
Volume 35 (2016), Number 3

Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts

Drew Skau[1] and Robert Kosara[1,2]

[1]UNC Charlotte [2]Tableau Research





Eurographics Conference on Visualization (EuroVis) 2016
R. Bujack, J. Kasten, and I. Viola
(Guest Editors)
Volume 35 (2016), Number 3

Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts

Drew Skau[1] and Robert Kosara[2,3]
[1]UNC Charlotte [2]Tableau Research

From "Show me the numbers" by Stephen C. Few

Straightening Tubular Flow for Side-by-Side Visualization

Paolo Angelelli, *Student Member, IEEE*, and Helwig Hauser, *Member, IEEE*

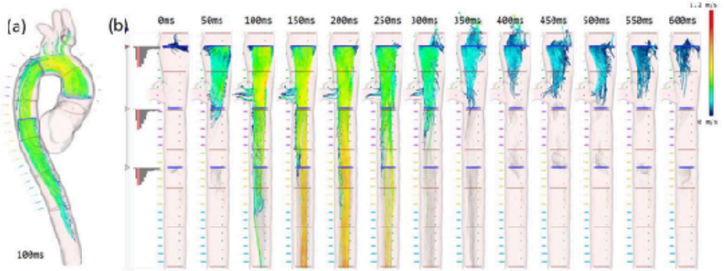


Fig. 1. (a) A timestep of an aortic flow dataset in its anatomical context, rendered using a conventional streamline visualization. (b) Slice-by-slice visualization of the straightened vector field, showing all the timesteps juxtaposed. The streamlines traced from the first seeding plane are rendered in focus, and the others in grey as context.

Visualizing statistical models: Removing the blindfold

Hadley Wickham, Dianne Cook and Heike Hofmann

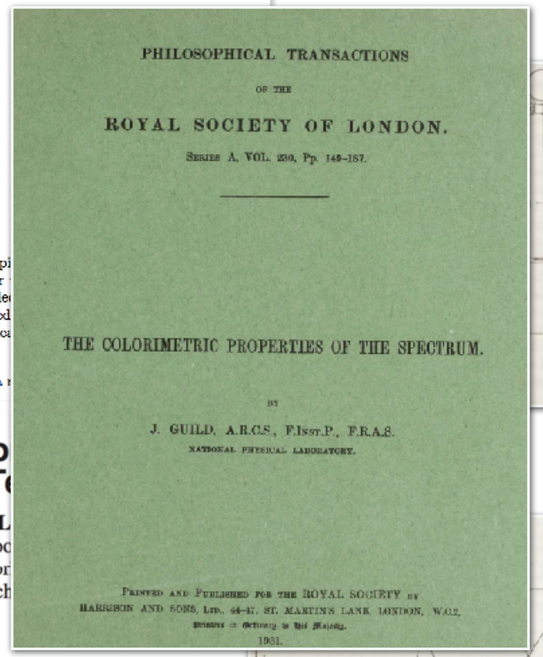
Department of Statistics MS-138
6100 Main St
Houston TX 77081
e-mail: hadley@rice.edu

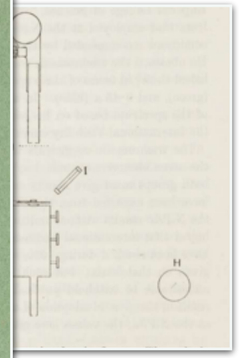
Department of Statistics
2415 Snedecor Hall
Ames IA 50011-1210
e-mail: dcook@iastate.edu

Department of Statistics
2413 Snedecor Hall
Ames IA 50011-1210
e-mail: hofmann@iastate.edu

Abstract: Visualization can help in model building, diagnosis, and in developi
about how a model summarizes data. This paper proposes three strategies for
models: (1) display the model in the data space, (2) look at all members of a colle
the process of model fitting, not just the end result. Each strategy is accompanied
ing MANOVA, classification algorithms, hierarchical clustering, ensembles of line
pursuit, self organizing maps and neural networks.

and phrases: model visualization, exploratory data analysis, data





Multi-Scale Banking to 45°

Jeffrey Heer and Maneesh Agrawala

Abstract—In his text *Visualizing Data*, William Cleveland demonstrates how the aspect ratio of a
perception of trends in the data. Cleveland proposes an optimization technique for computing
average absolute orientation of line segments in the chart is equal to 45 degrees. This techn
designed to maximize the discriminability of the orientations of the line segments in the chart. In t
result and describe two new extensions. First, we propose alternate optimization criteria designe
perception of line segment orientations. Second, we develop *multi-scale banking*, a technique t
with banking to 45°. Our technique automatically identifies trends at various frequency scales and
for each of these scales. We demonstrate the utility of our techniques in a range of visualization to

Index Terms—Information visualization, banking to 45 degrees, line charts, time-series, sparklines,

How Deceptive are Deceptive Visualizatio
An Empirical Analysis of Common Distortion Te

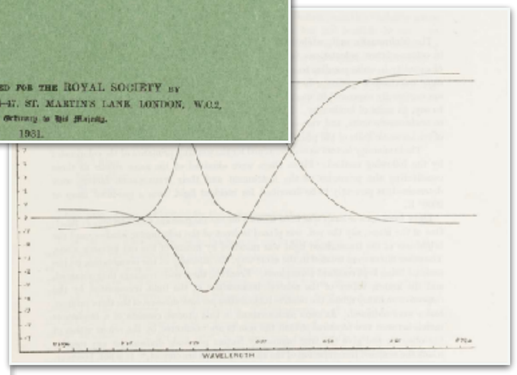
Anshul Vikram Pandey
School of Engineering,
New York University
anshul.pandey@nyu.edu

Katharina Rall
School of Law,
New York University
kr1326@nyu.edu

Margaret L
Schoo
New Yor
satterth@exch

Oded Nov
School of Engineering,
New York University
onov@nyu.edu

Enrico Bertini
School of Engineering,
New York University
enrico.bertini@nyu.edu



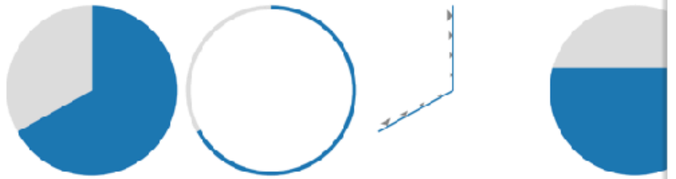
ABSTRACT
In this paper, we present an empirical analysis of deceptive visualizations. We start with an in-depth analysis of what deception means in the context of data visualization, and categorize deceptive visualizations based on the type of deception they lead to. We identify popular distortion techniques and the type of visualizations those distortions can be applied to, journalism [14, 35], specialists and laypersons are using data to shape compelling, informative, and convincing narratives, conveyed through or supported by visualizations. While the use of such visual depictions as persuasion devices is not new, the popular use of visualizations has undoubtedly increased due in part to user-friendly software that allows non-experts to create visualizations. As such practices become

Eurographics Conferen
K.-L. Ma, G. Santucci,
(Guest Editors)

Arcs, Angles, or Areas:
Individual Data Encodings in Pie and Donut Charts

Drew Skau[1] and Robert Kosara[1,2]

[1]UNC Charlotte [2]Tableau Research



High-Speed Visual Estimation Using Preattentive Processing

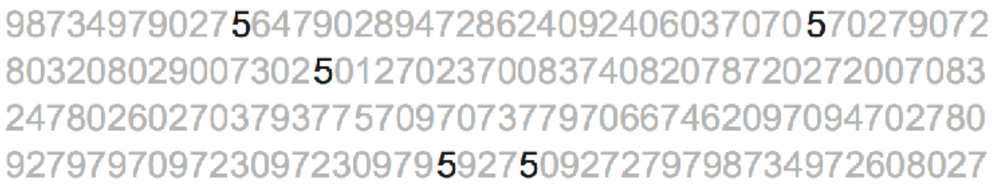
CHRISTOPHER G. HEALEY, KELLOGG S. BOOTH, and JAMES T. ENNS
The University of British Columbia

A new method is presented for performing rapid and accurate numerical estimation. The method is derived from an area of human cognitive psychology called preattentive process-

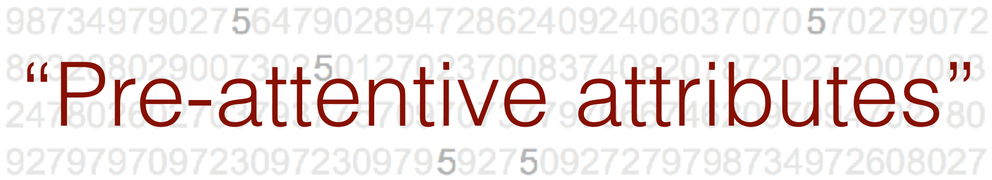
Three most common data visualization mistakes

Attitude

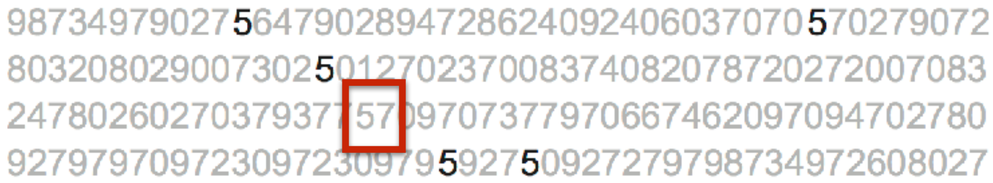
98734979027564790289472862409240603707057027907280320802900730250127023700837408207872027200708324780260270379377570970737797066746209709470278092797970972309723097959275092727979873497260802 7



From "Show me the numbers" by Stephen C. Few

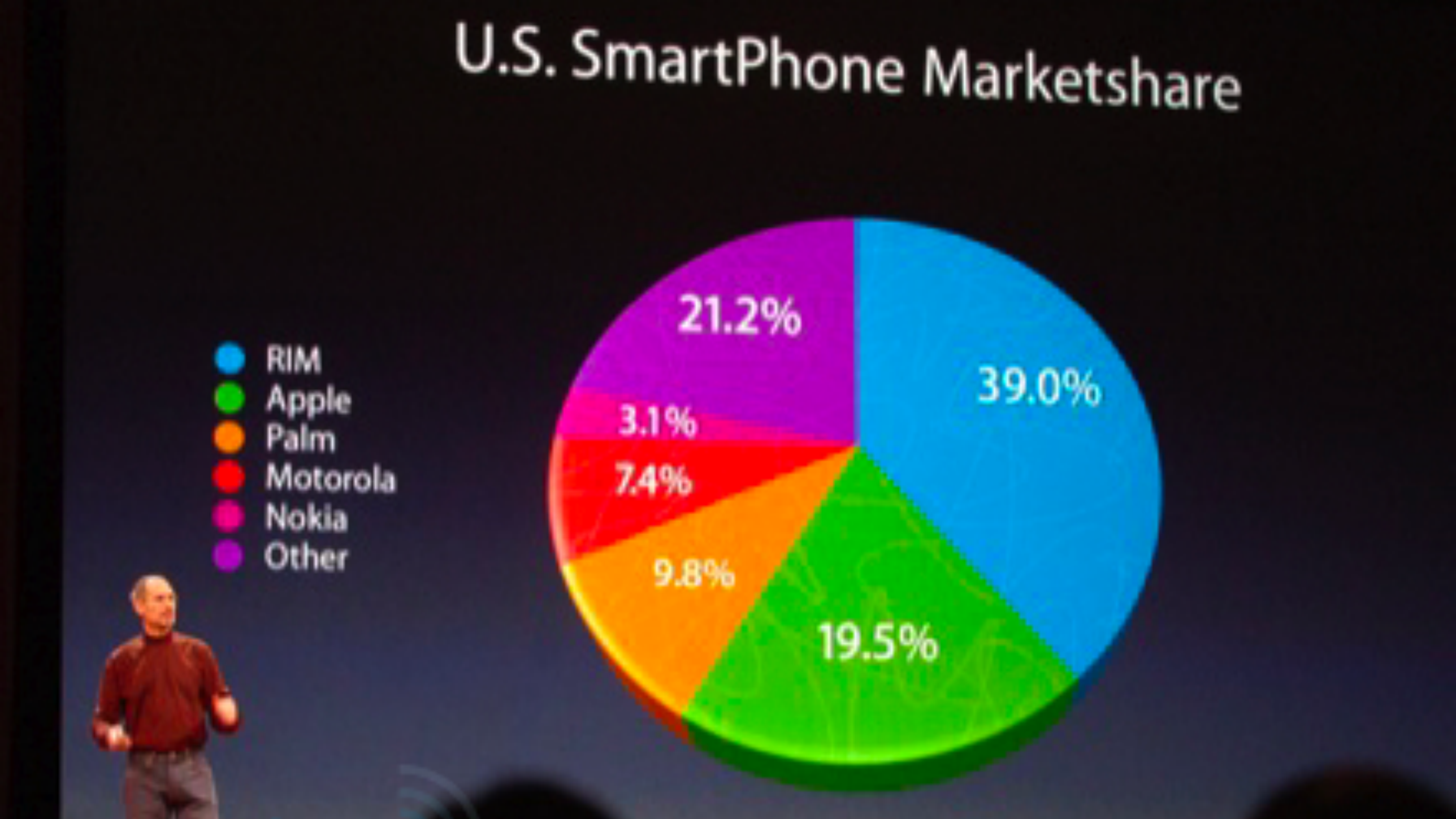


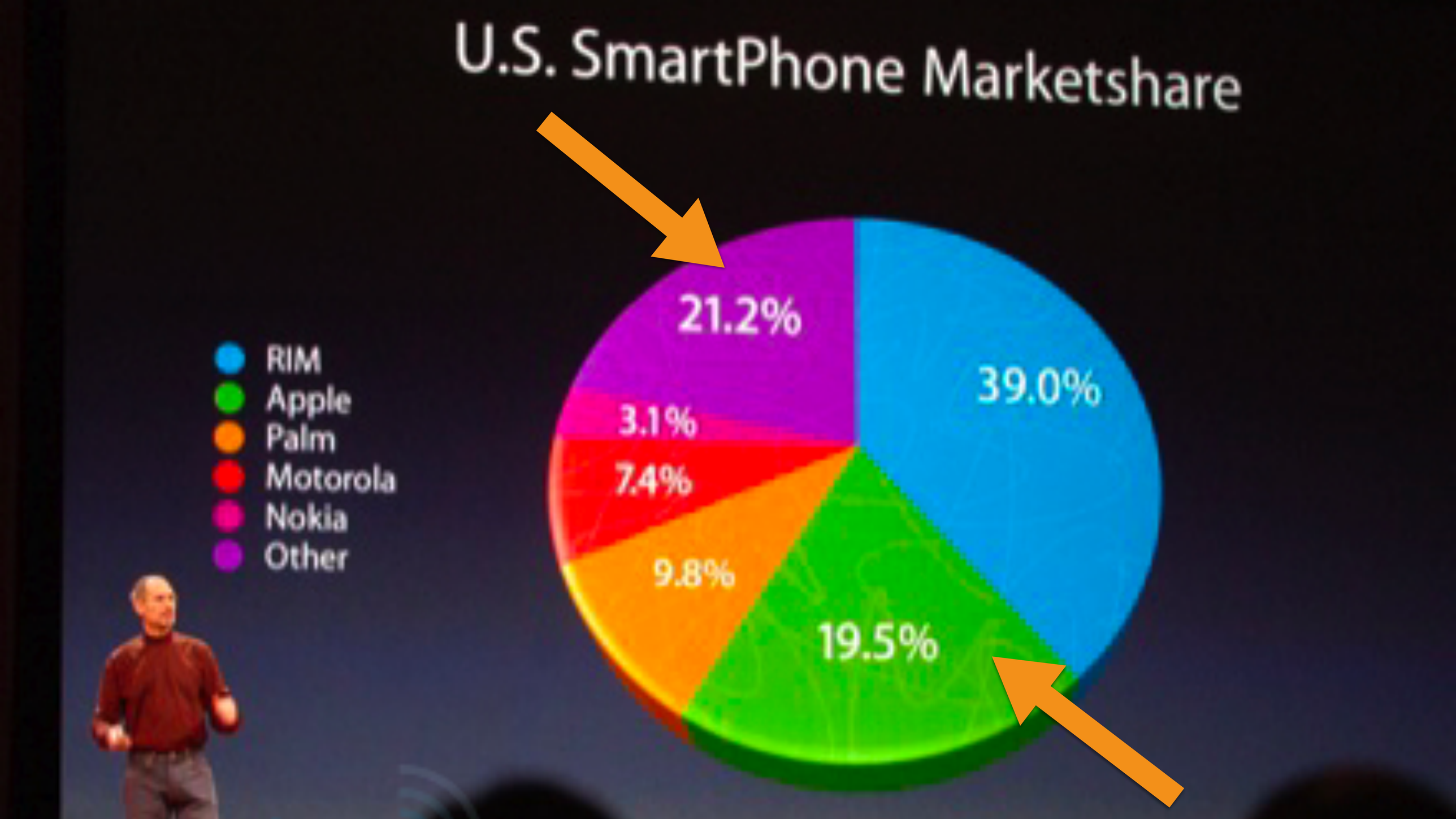
From "Show me the numbers” by Stephen C. Few



From "Show me the numbers" by Stephen C. Few

U.S. SmartPhone Marketshare




U.S. SmartPhone Marketshare
RIM
Apple
Palm
Motorola
Nokia
Other
21.2%
39.0%
3.1%
7.4%
9.8%
19.5%

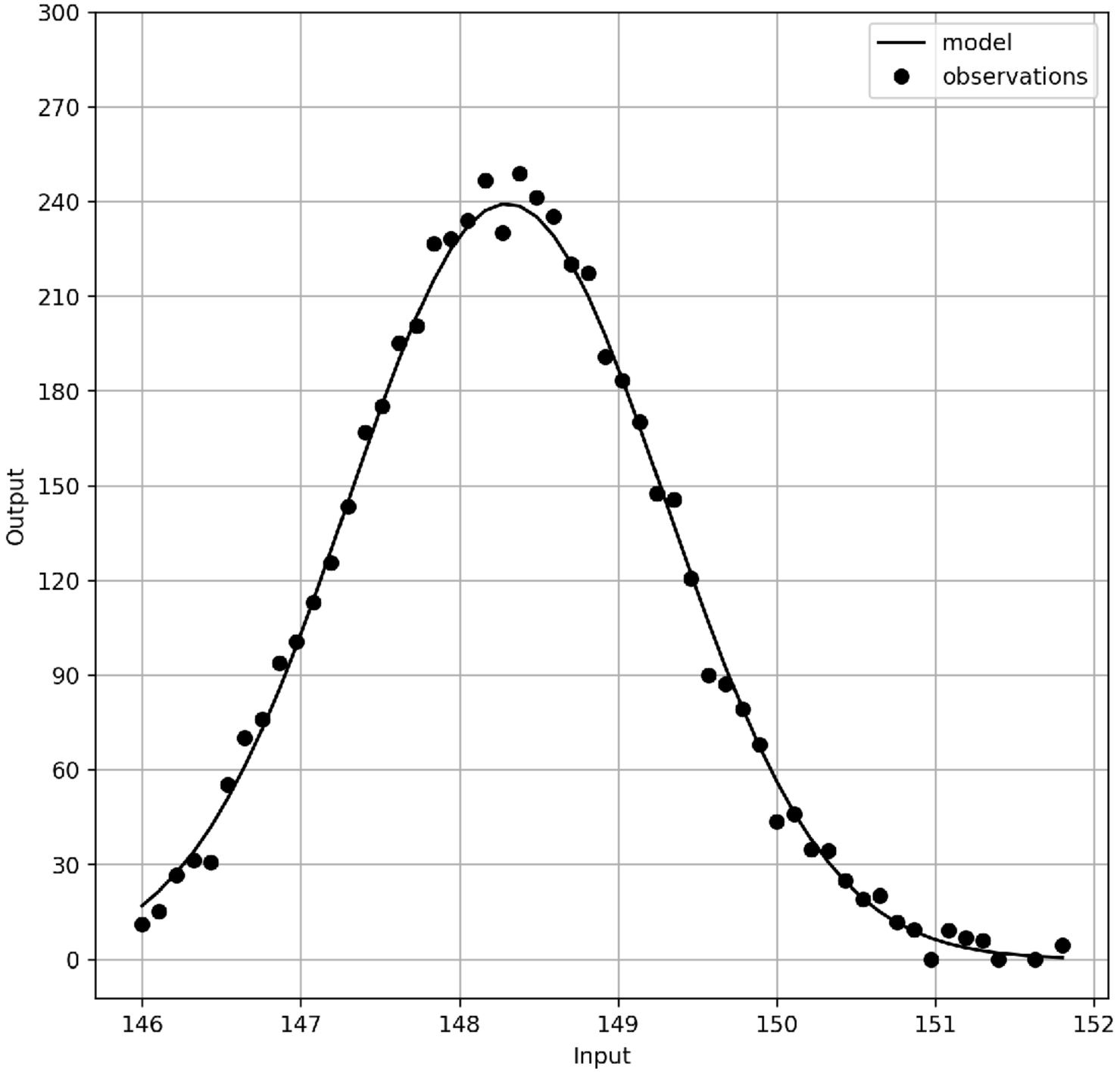
Three most common data visualization mistakes

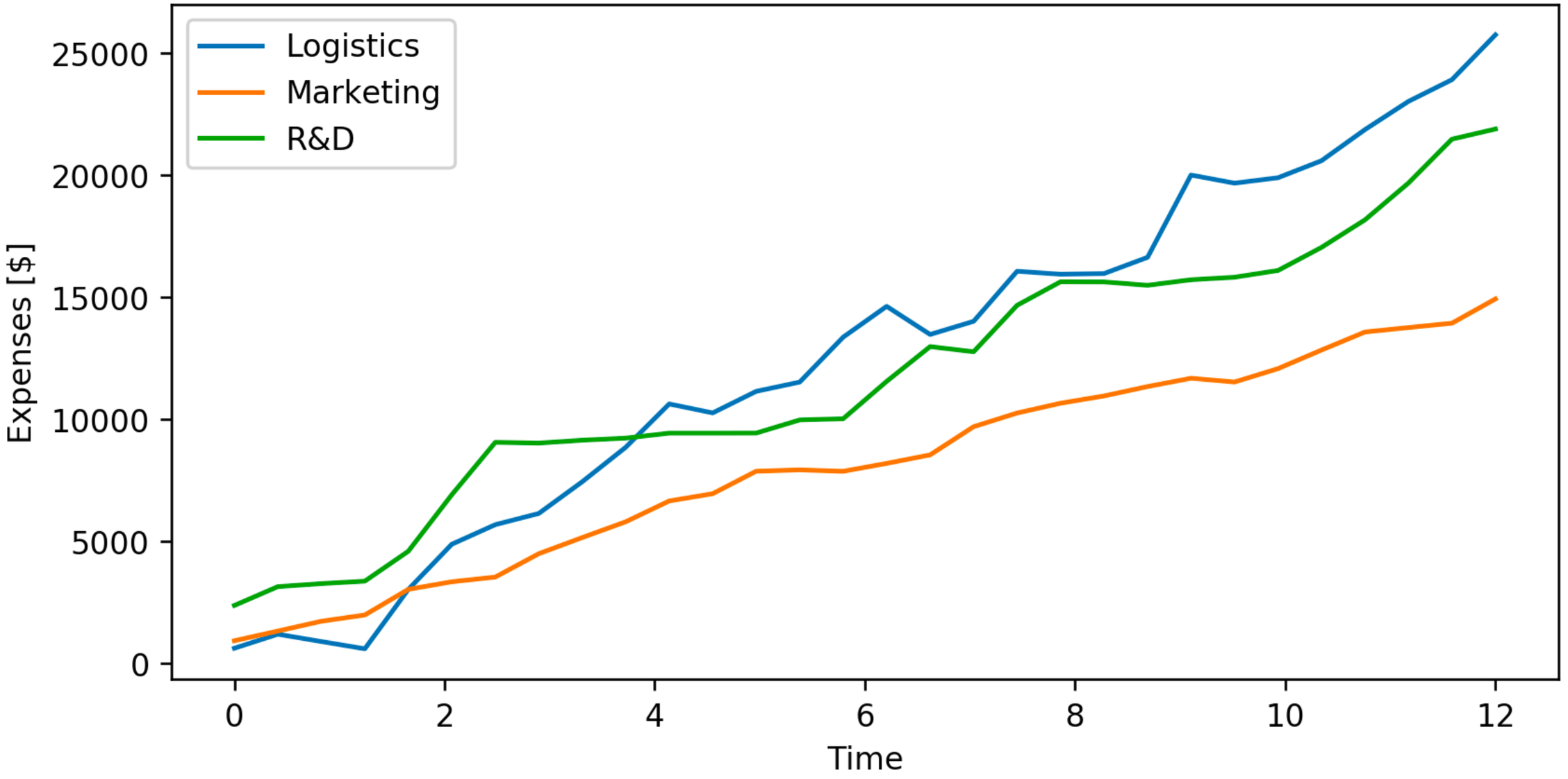
Attitude

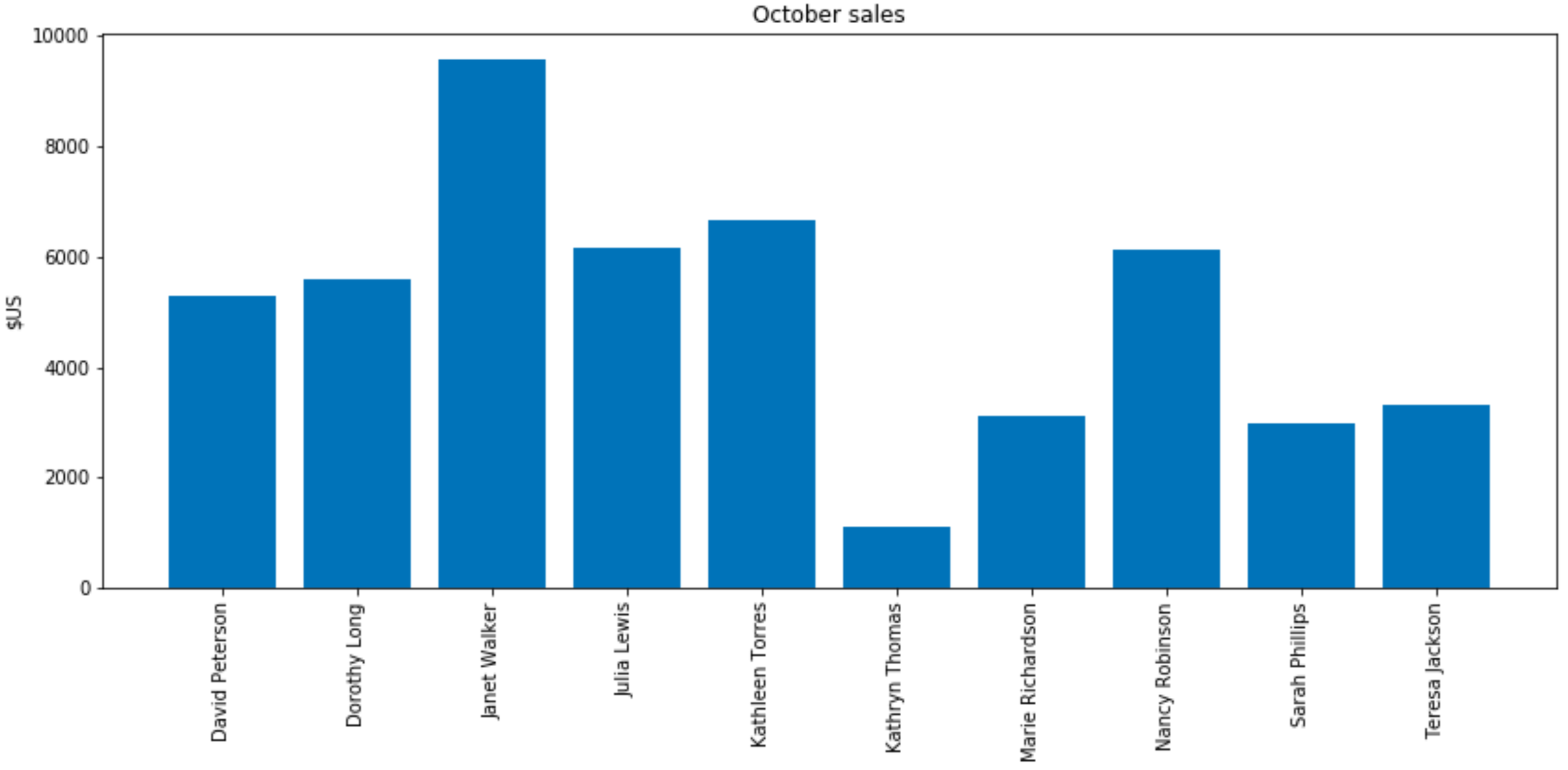
Explanatory vs. Exploratory. Evidence-based.
Someone will lie to you

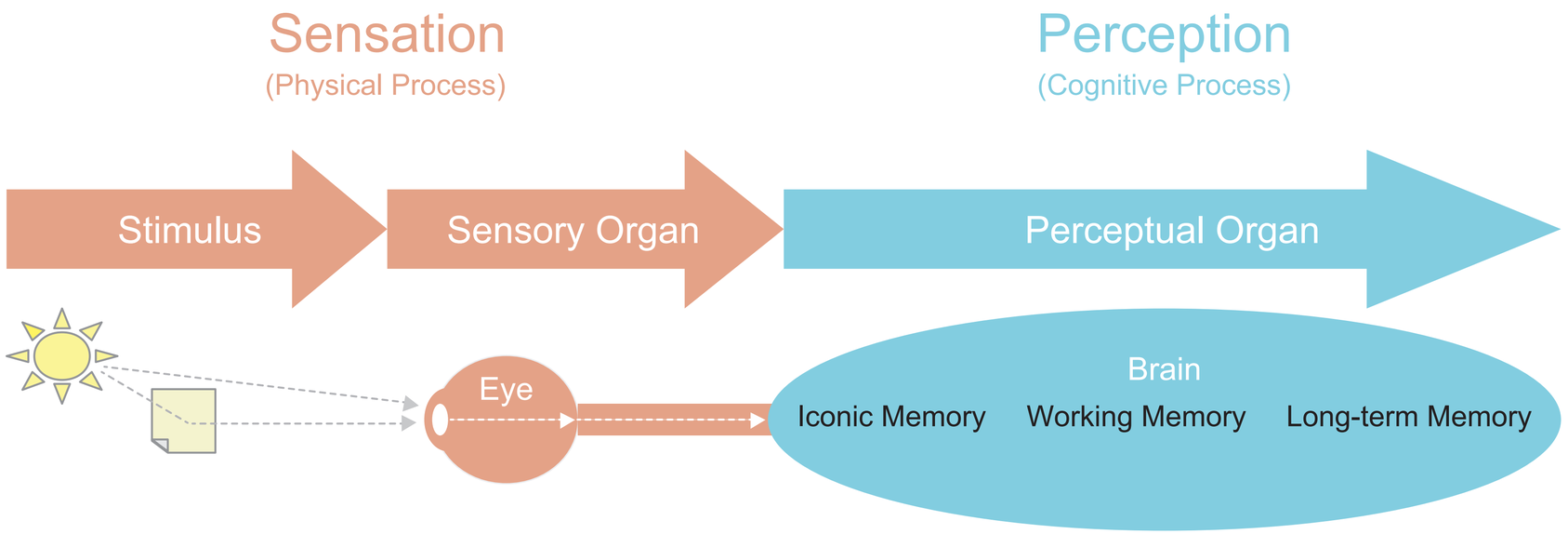
Three most common data visualization mistakes

2

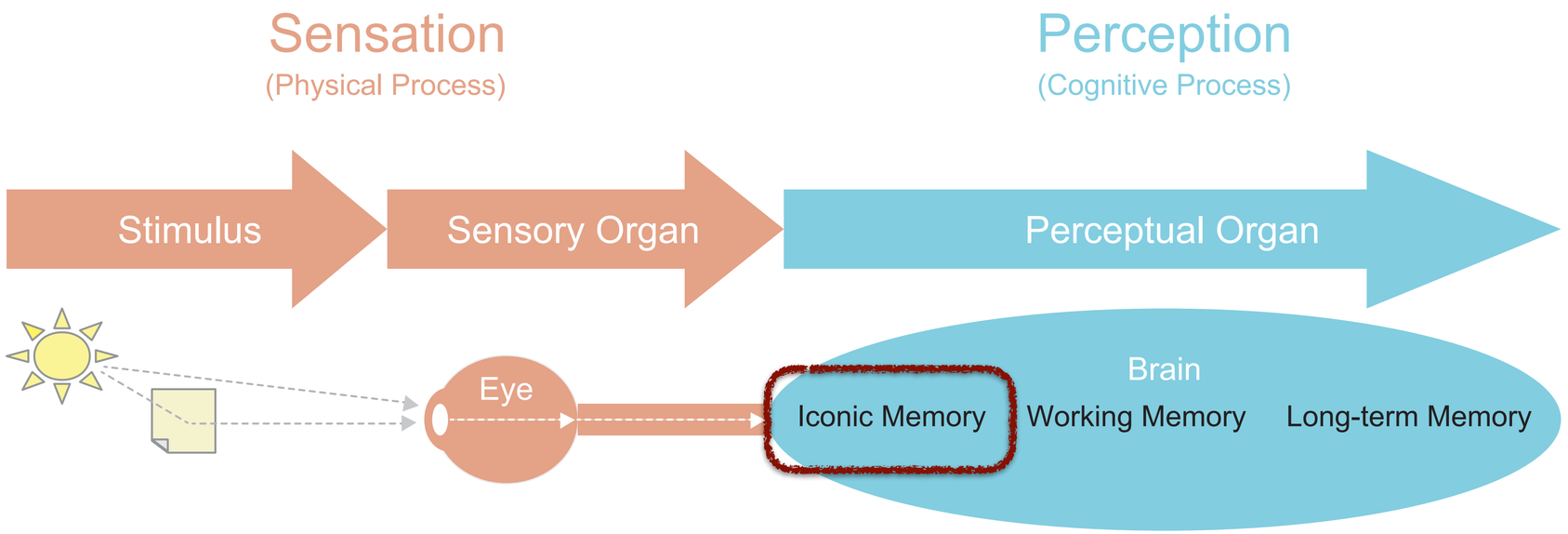

model
observations
Output
Input
300
270
240
210
180
150
120
90
60
30
0
146
147
148
149
150
151
152


Logistics
Marketing
R&D
25000
20000
15000
10000
5000
0
Expenses [$]
0
2
4
6
8
10
12
Time

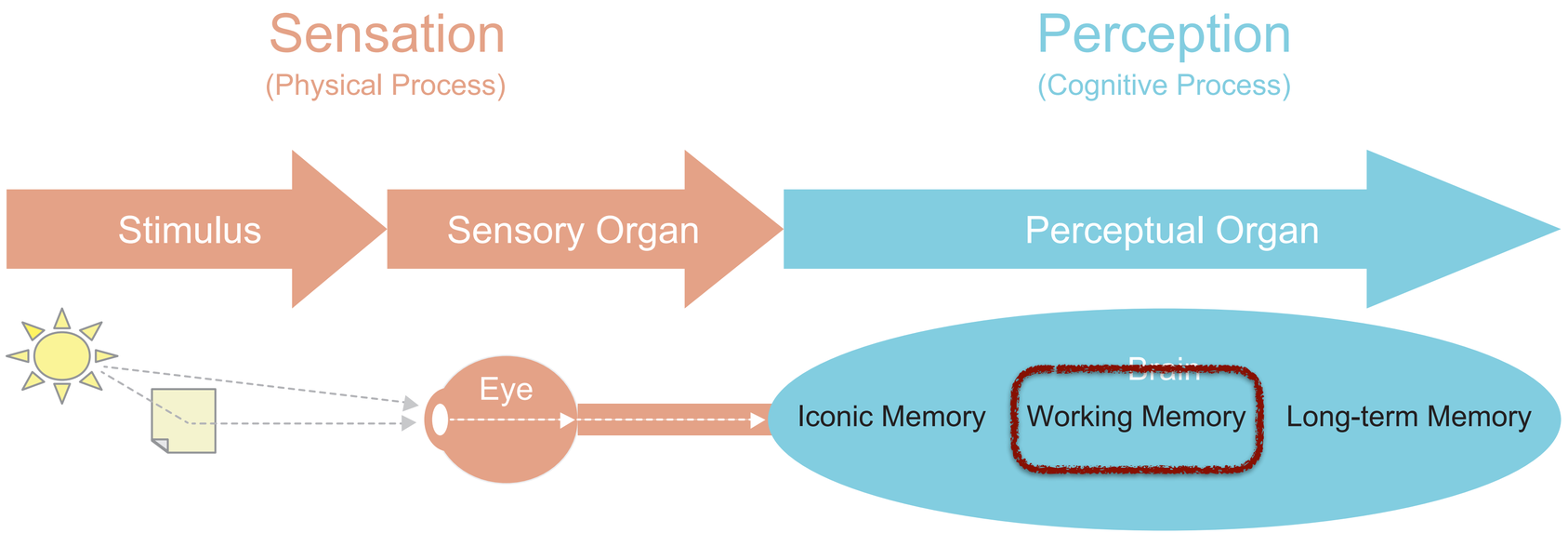

October sales
$US
10000
8000
6000
4000
2000
0
David Peterson
Dorothy Long
Janet Walker
Julia Lewis
Kathleen Torres
Kathryn Thomas
Marie Richardson
Nancy Robinson
Sarah Phillips
Teresa Jackson



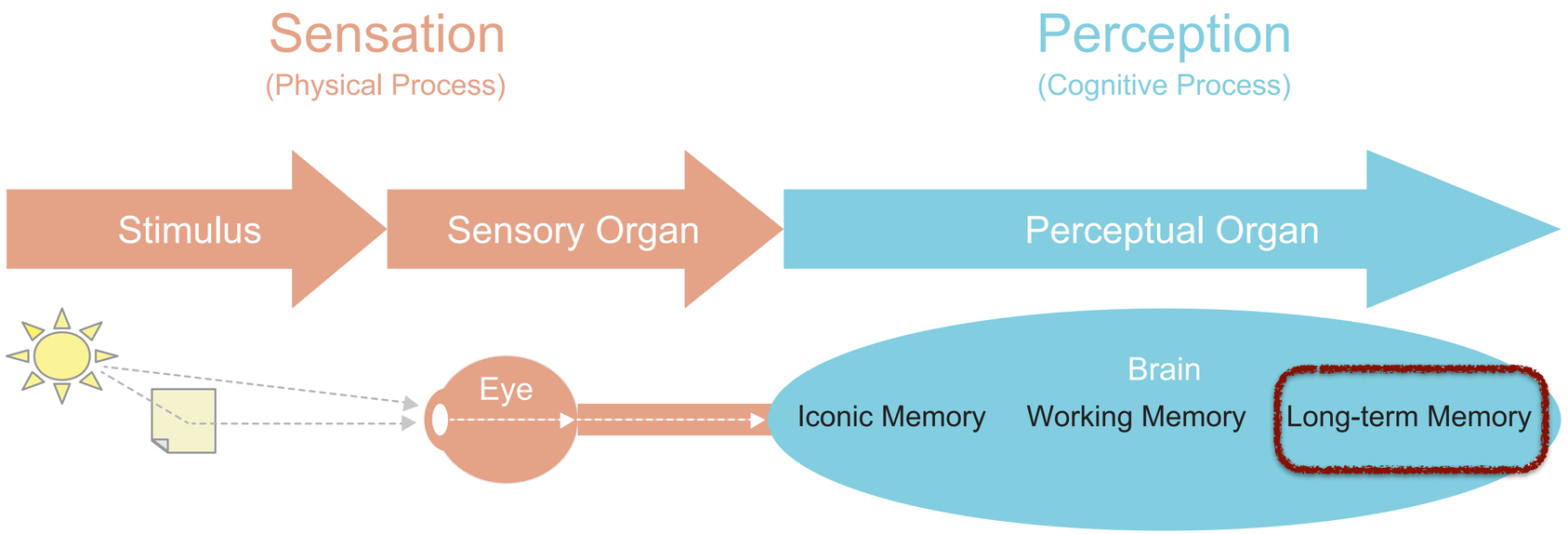
From "Show me the numbers" by Stephen C. Few



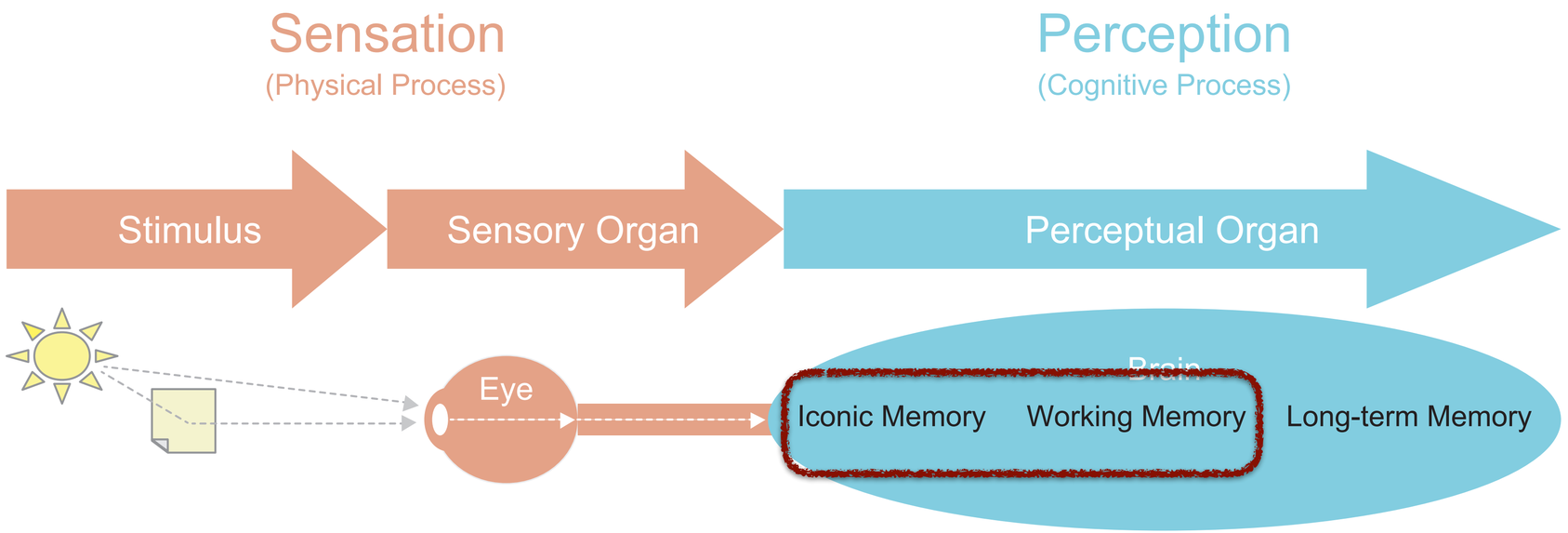
From "Show me the numbers" by Stephen C. Few



From "Show me the numbers" by Stephen C. Few

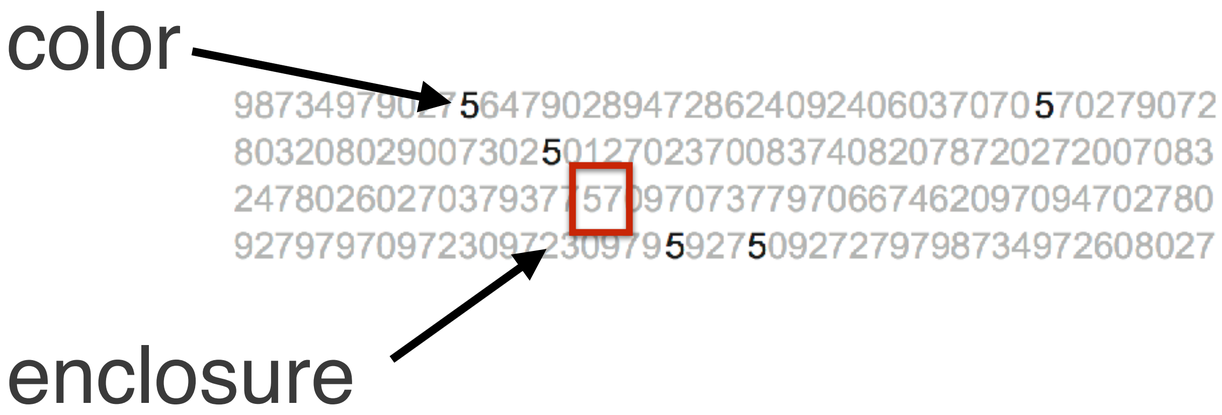


From "Show me the numbers" by Stephen C. Few



“Pre-attentive attributes”

From "Show me the numbers” by Stephen C. Few



From "Show me the numbers" by Stephen C. Few

NEUROSCIENCE

You Now Have a Shorter Attention Span Than a Goldfish

Kevin McSpadden
May 14, 2015

For more, visit TIME Health.

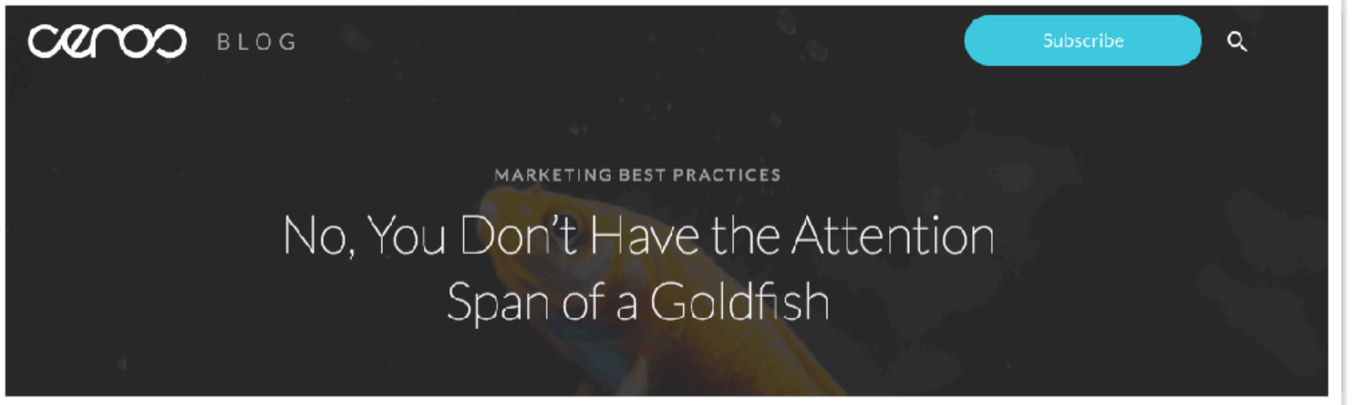
NEUROSCIENCE

You Now Have a Shorter Attention Span Than a Goldfish

Kevin McSpadden
May 14, 2015

TIME Health

For more, visit TIME Health.



The Effect of Aesthetic on the Usability of Data Visualization

Nick Cawthon and Andrew Vande Moere
The University of Sydney, Australia
nick@student.usyd.edu.au, andrew@arch.usyd.edu.au

w Have a Shorter Attention Span Goldfish

th.

Behaviour & Information Technology, Vol. 25, No. 2, March-April 2006, 115–126

Taylor & Francis
Taylor & Francis Group

Attention web designers: You have 50 milliseconds to make a good first impression!

GITTE LINDGAARD†*, GARY FERNANDES‡, CATHY DUDEK§ and J. BROWÑ¶

Subscribe

MARKETING BEST PRACTICES

No, You Don't Have the Attention Span of a Goldfish

The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments

Alexandre N. Tuch, [a,*] Eva E. Presslaber, [a] Markus Stöcklin, [a] Klaus Opwis, [a] Javier A. Bargas-Avila [b]

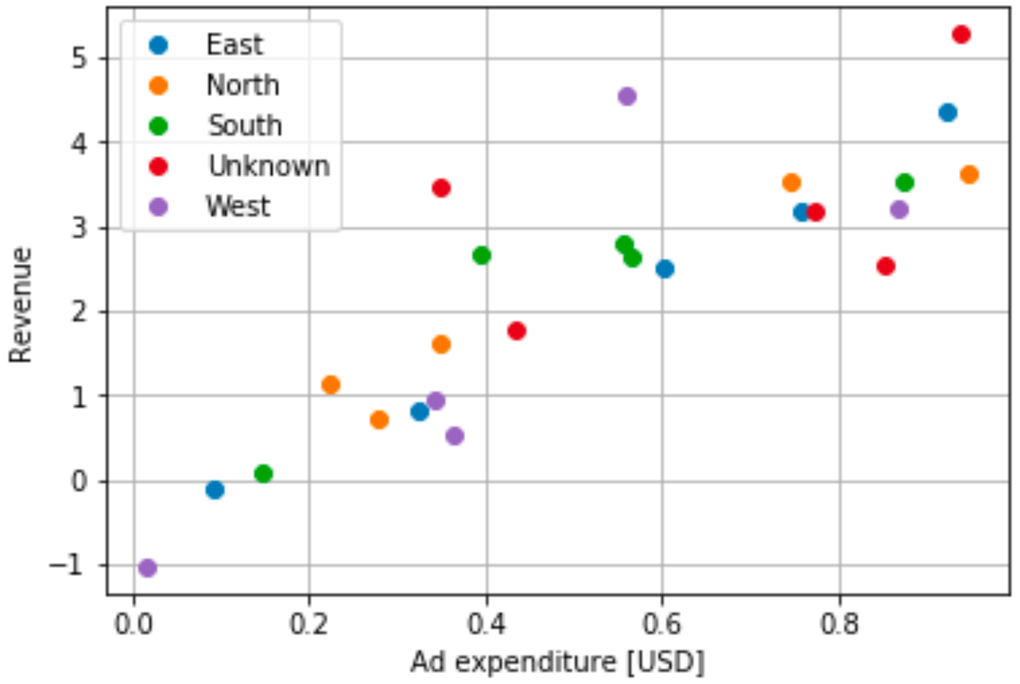
Edward Tufte
http://edwardtufte.com

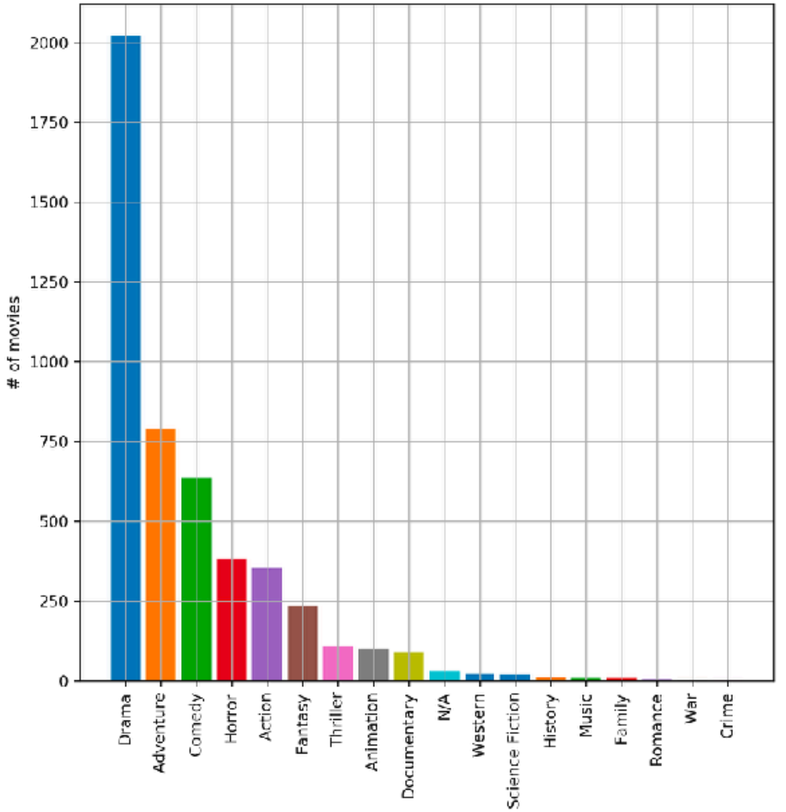
Data-ink ratio

(signal-to-noise ratio)

Above all, show data

No data — no ink


East
North
South
Unknown
West
Revenue
Ad expenditure [USD]
5
4
3
2
1
0
−1
0.0
0.2
0.4
0.6
0.8


# of movies
2000
1750
1500
1250
1000
750
500
250
0
Drama
Adventure
Comedy
Horror
Action
Fantasy
Thriller
Animation
Documentary
N/A
Western
Science Fiction
History
Music
Family
Romance
War
Crime

Signal

Noise

Signal

Noise

Data

Noise

Signal	Noise
Data	Noise
Useful information	Useless information

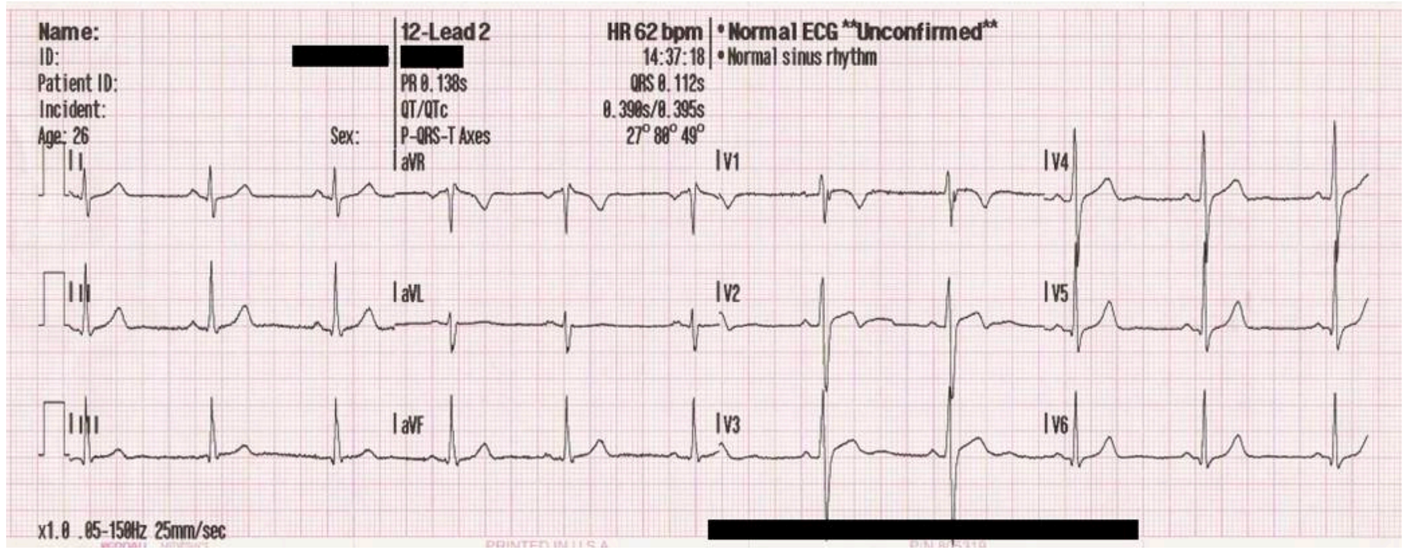


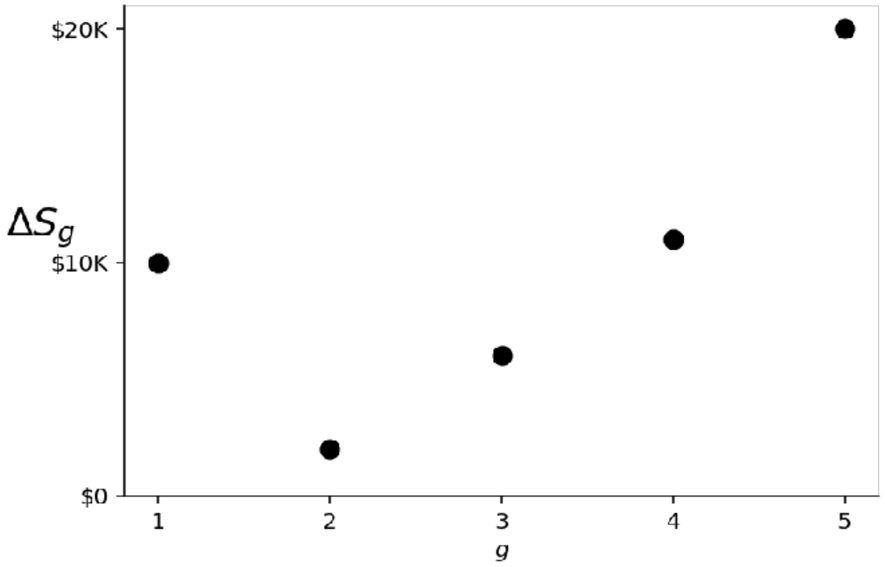




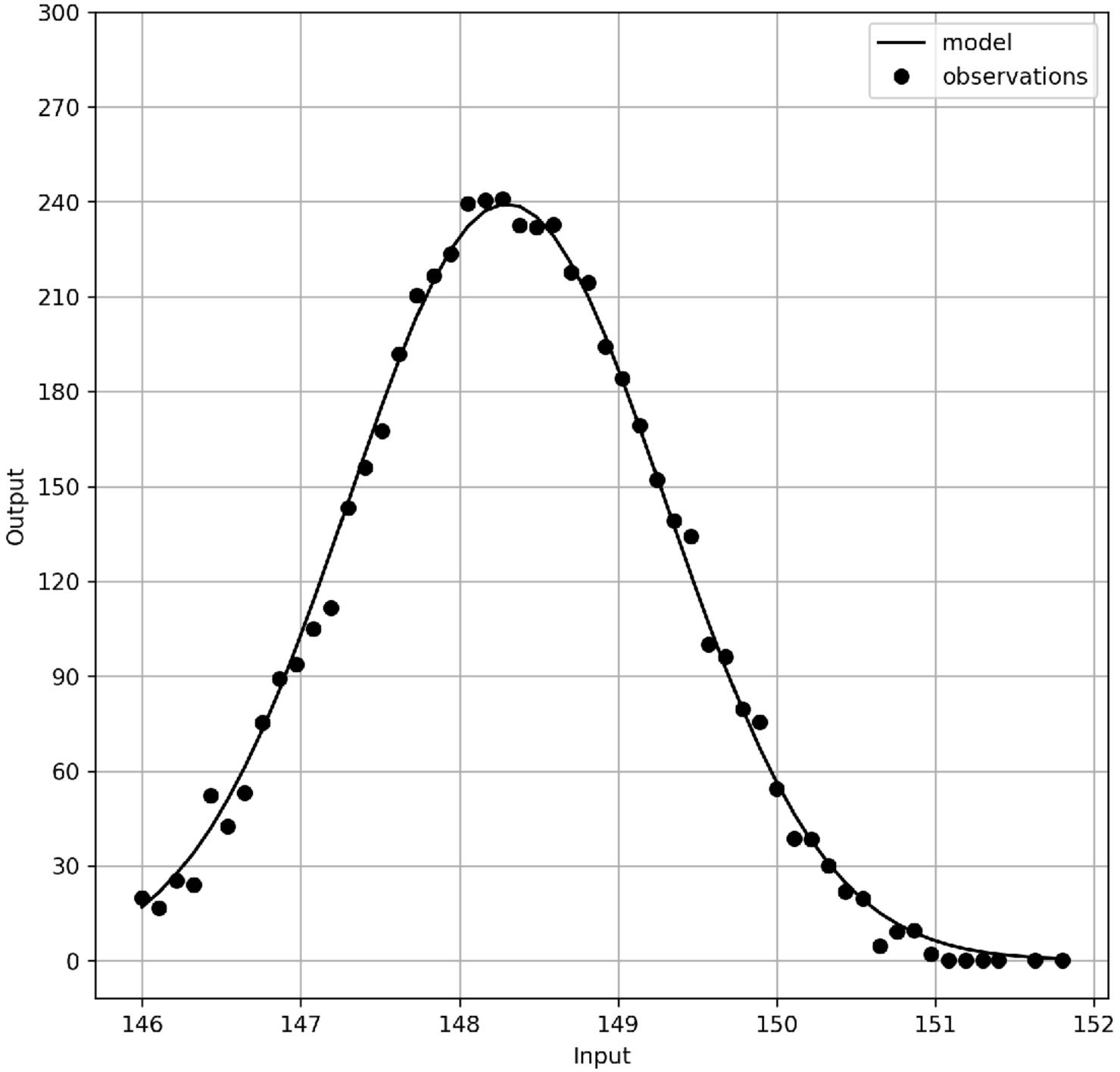


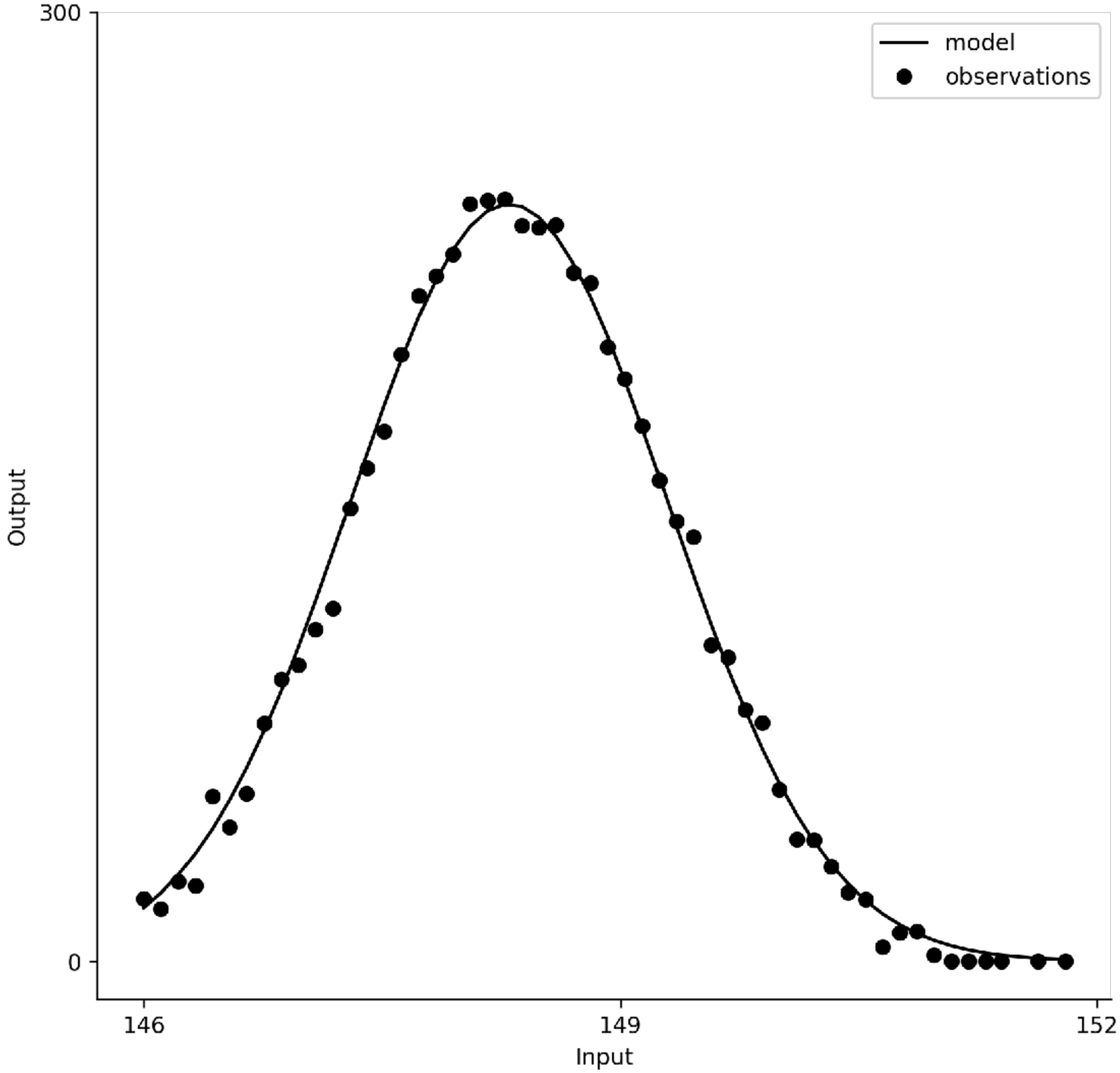


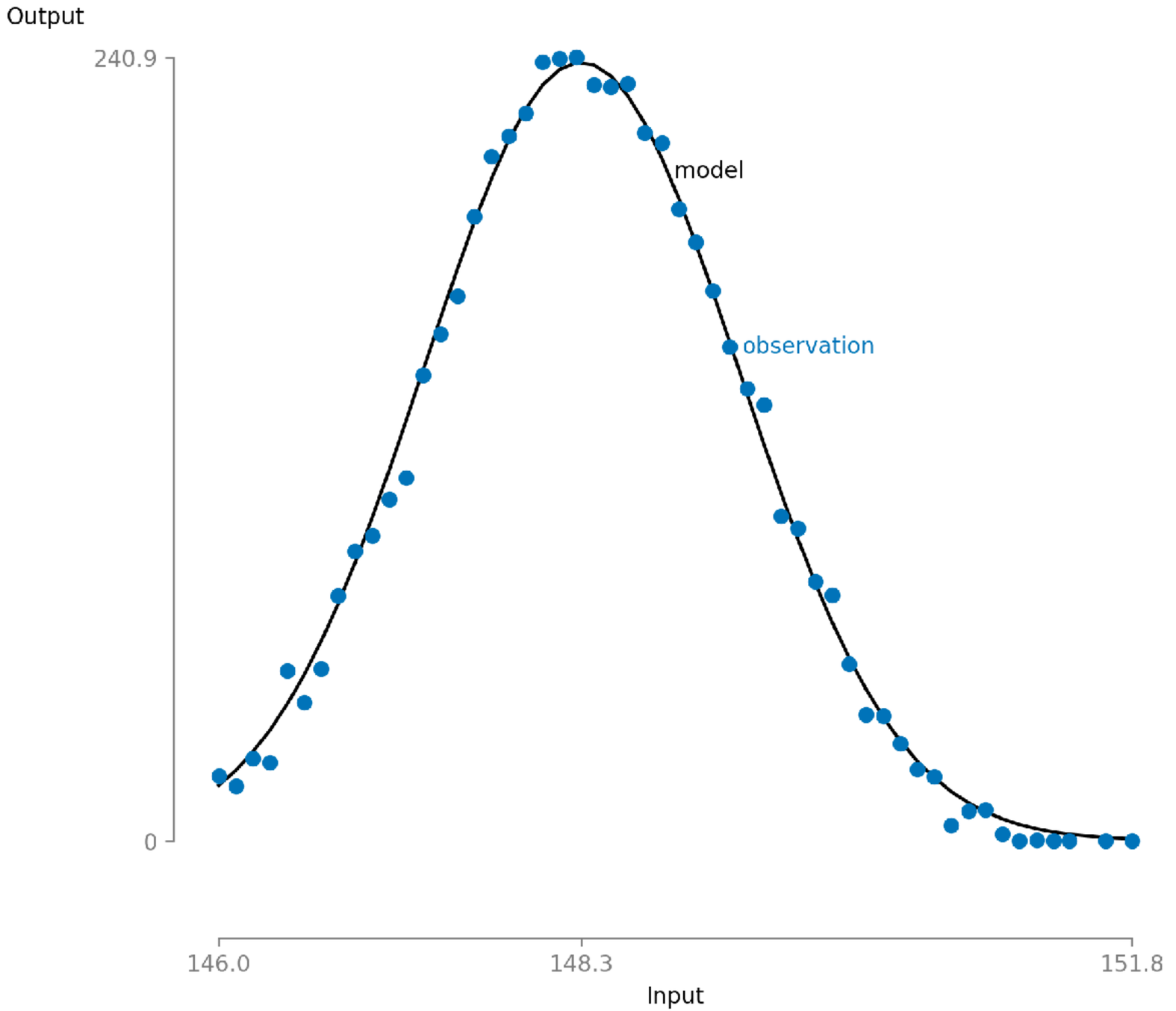


Name:
ID:
Patient ID:
Incident:
Age: 26
Sex:
12-Lead 2
PR 0.138s
QT/QTc
P-QRS-T Axes
aVR
HR 62 bpm
14:37:18
QRS 0.112s
0.390s/0.395s
27° 88° 49°
• Normal ECG **Unconfirmed**
• Normal sinus rhythm
I
II
III
aVL
aVF
V1
V2
V3
V4
V5
V6
x1.0 .05-150Hz 25mm/sec



From "Show me the numbers" by Stephen C. Few


model
observations
Output
Input
300
270
240
210
180
150
120
90
60
30
0
146
147
148
149
150
151
152

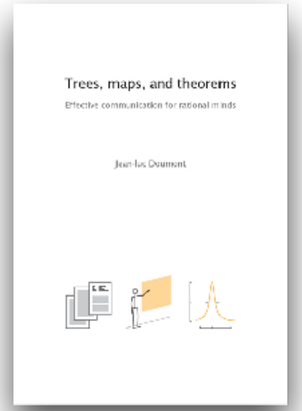

300
0
Output
146
149
152
Input
model
observations

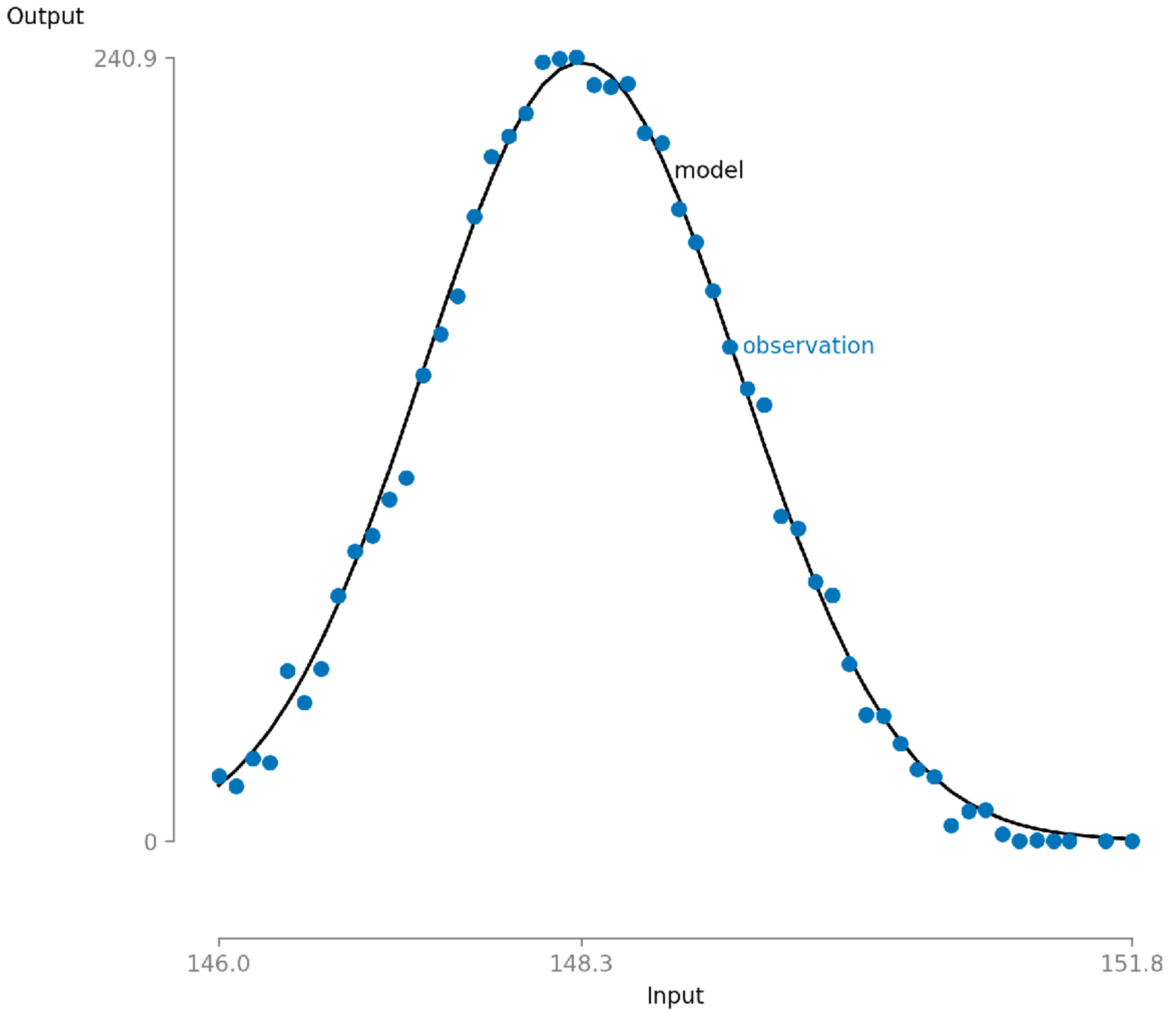

Output
240.9
0
model
observation
146.0
148.3
151.8
Input

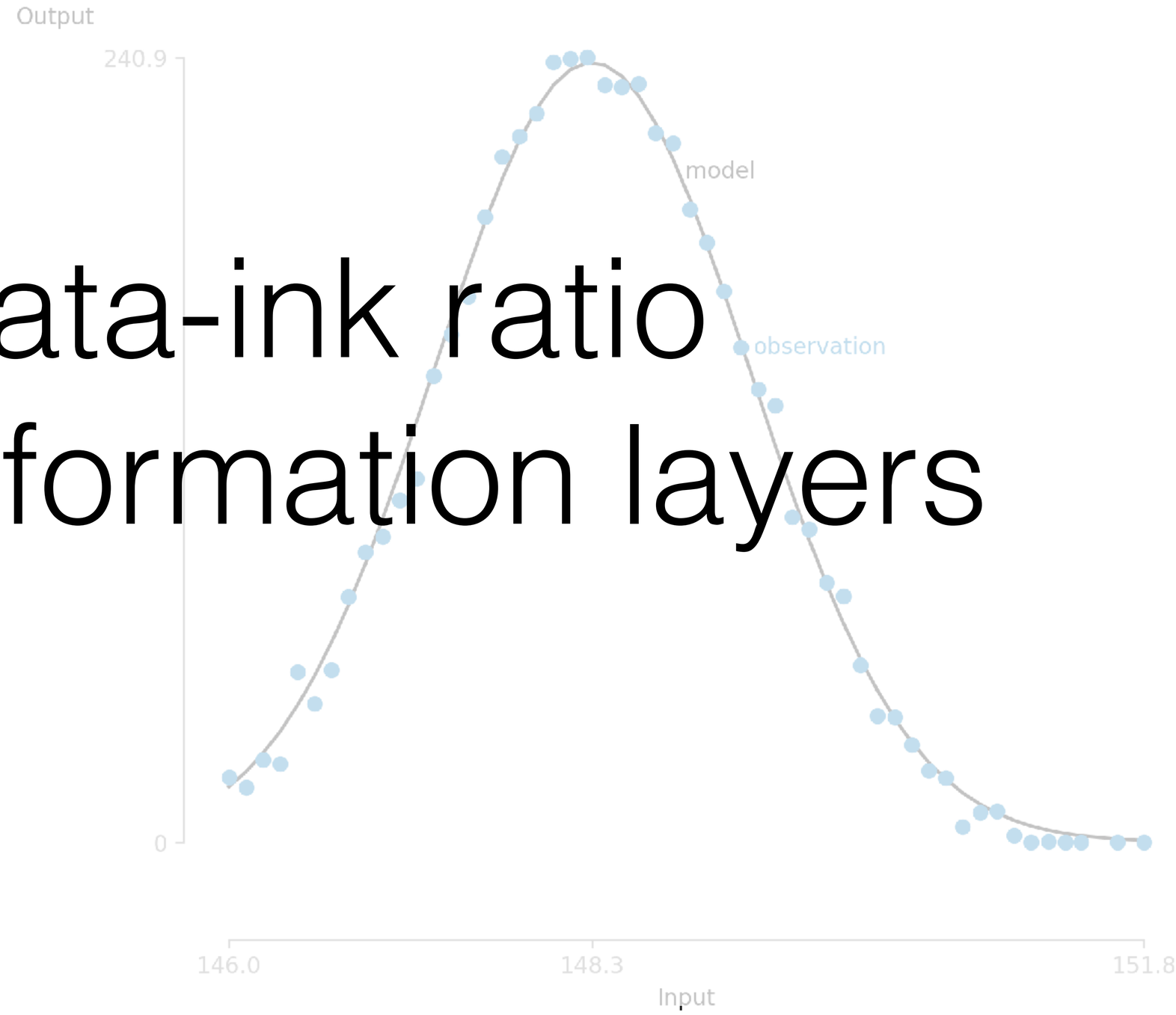
Jean-luc Doumont
http://www.principiae.be

Information layers

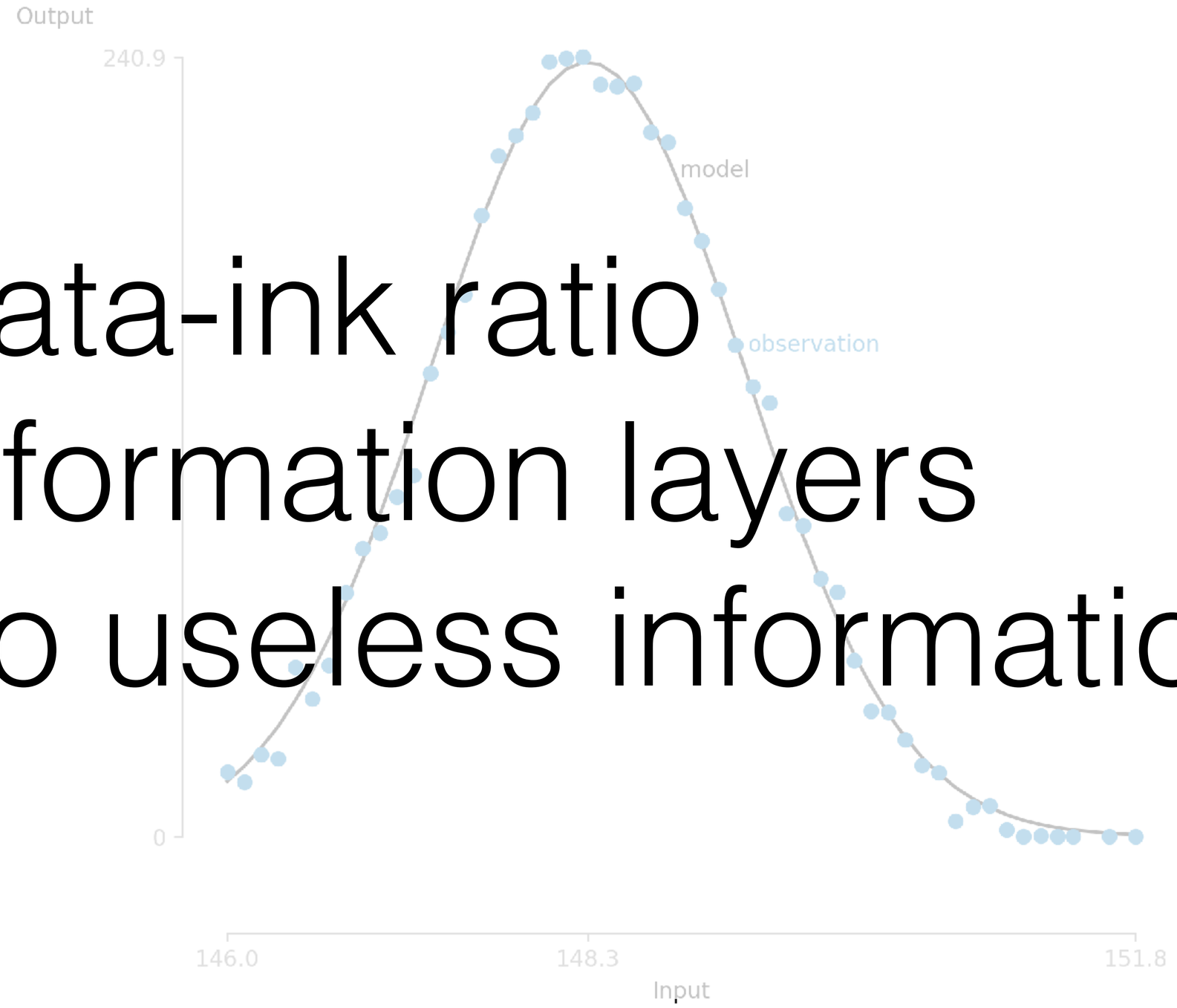
(Useful redundancy)



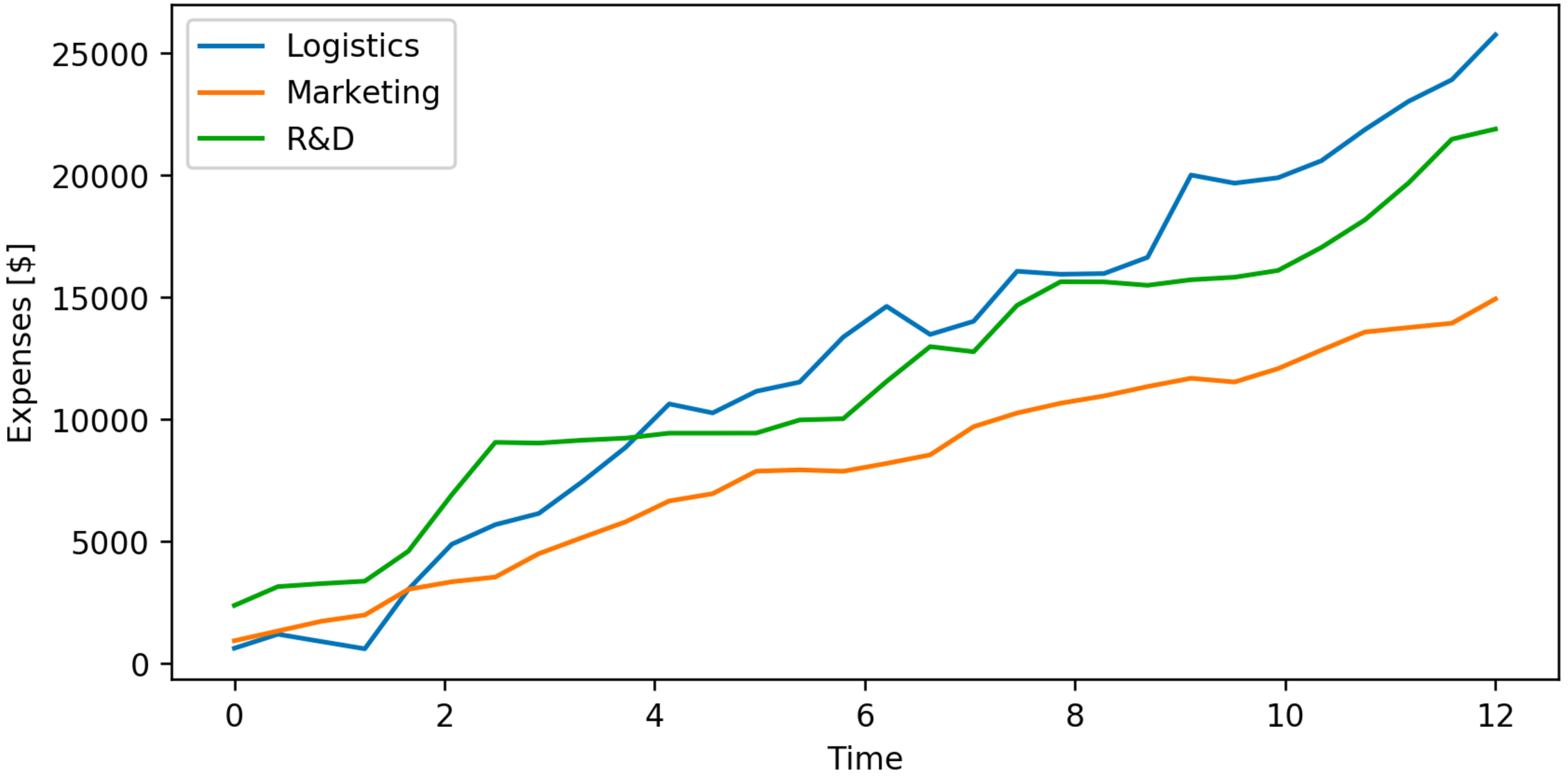

Output
240.9
0
model
observation
146.0
148.3
151.8
Input

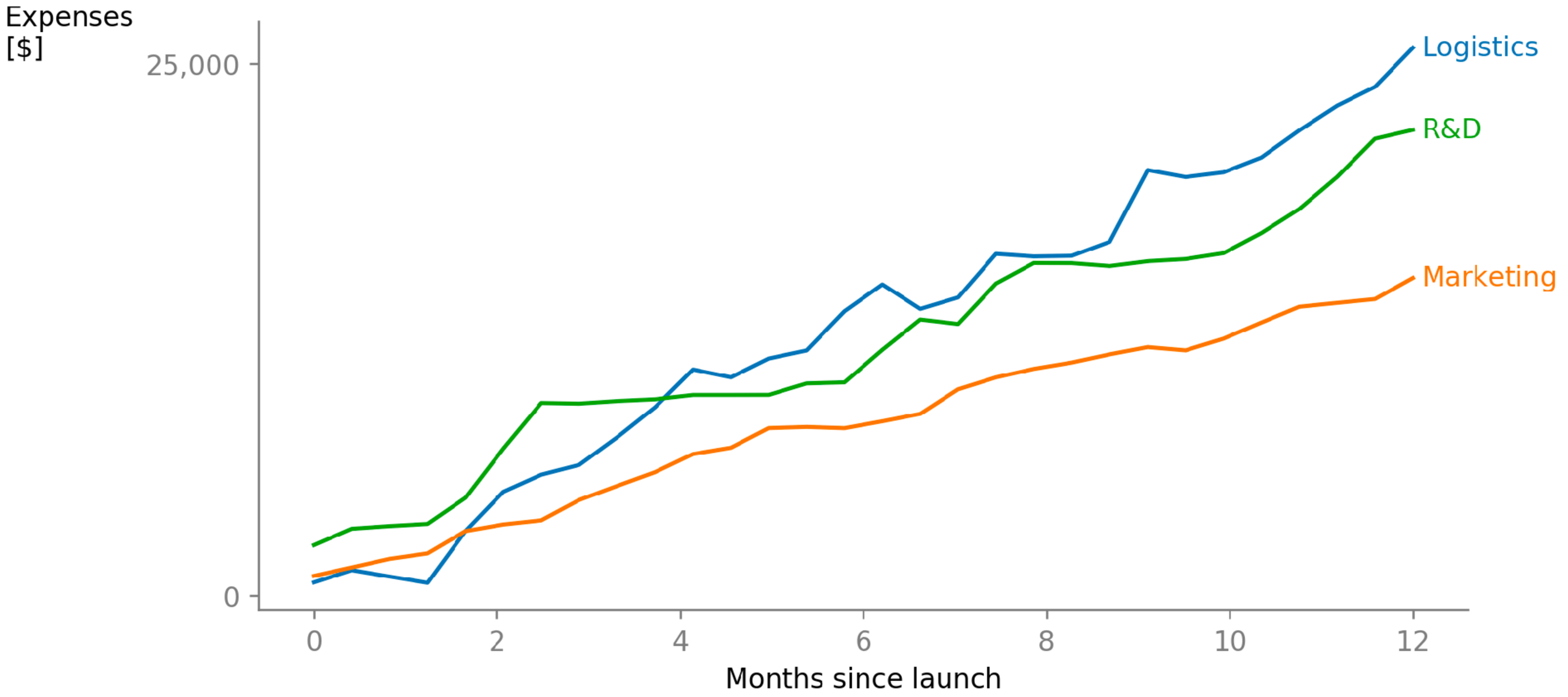


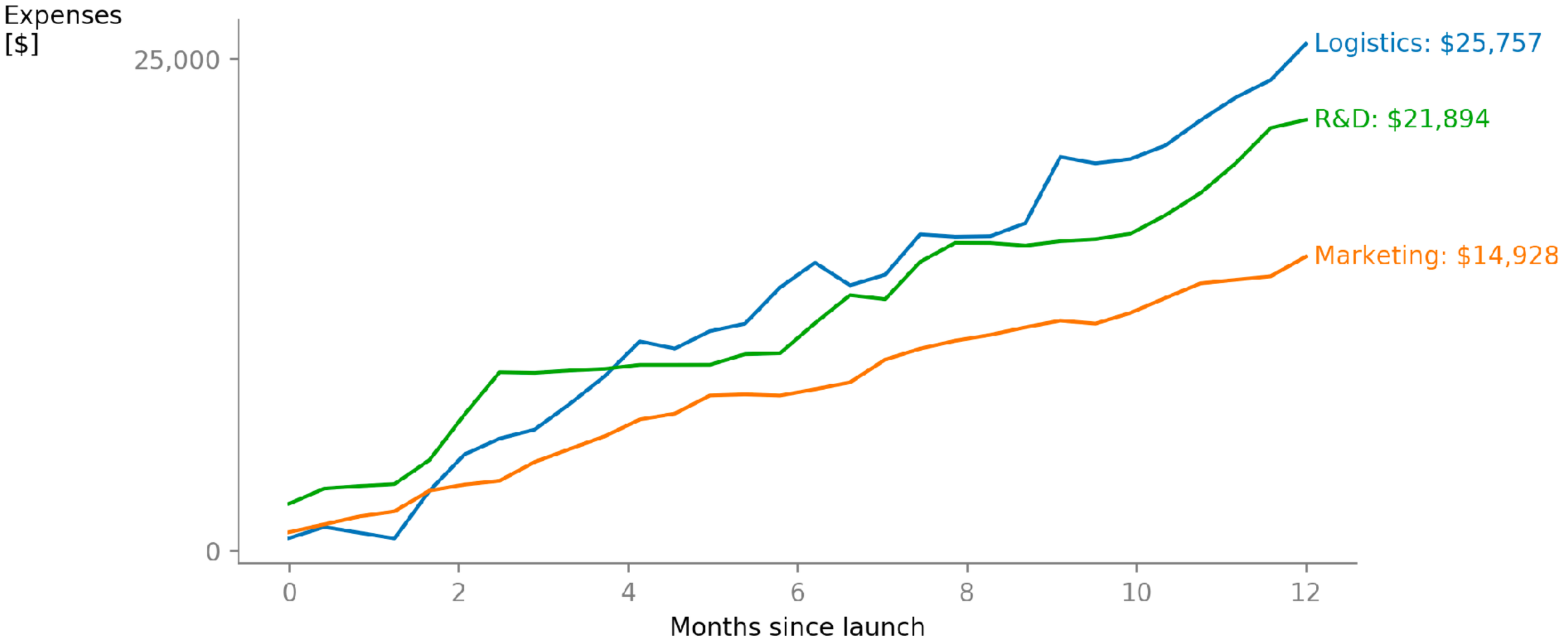
Data-ink ratio
Information layers

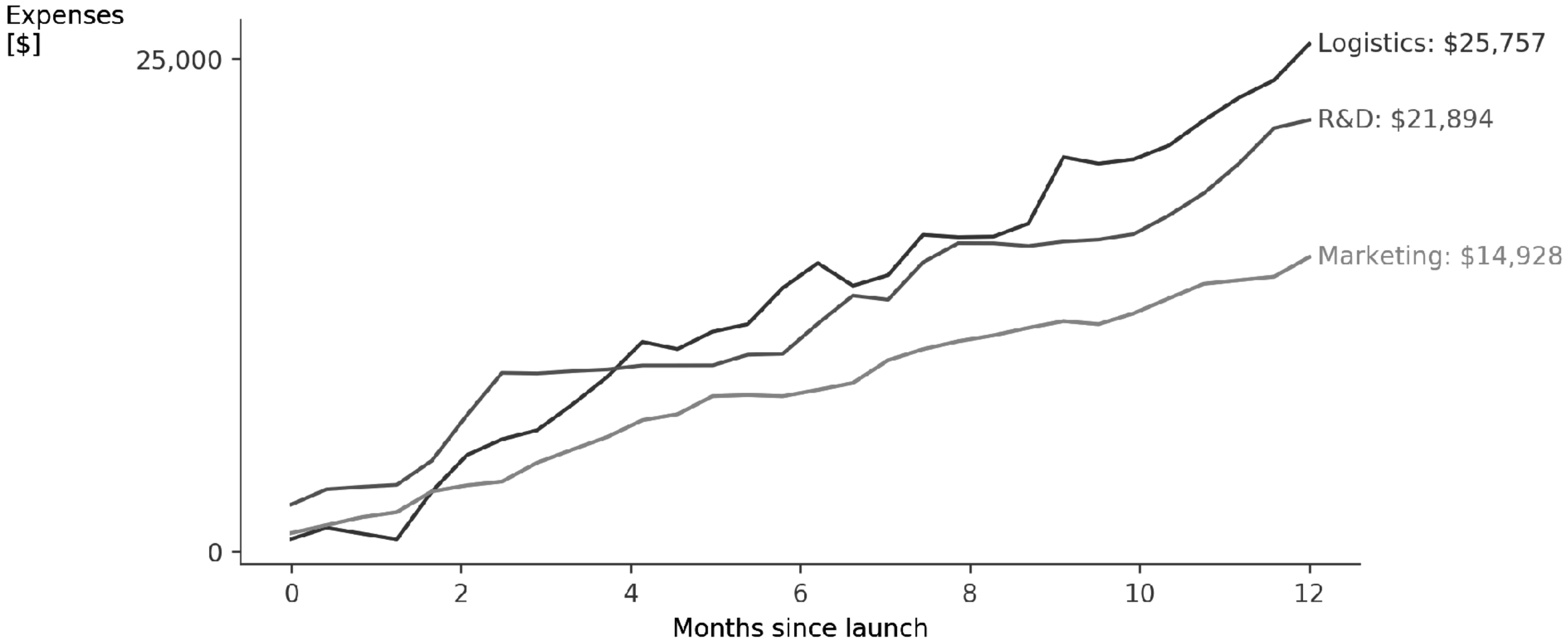


Data-ink ratio
Information layers
No useless information


Logistics
Marketing
R&D
25000
20000
15000
10000
5000
0
Expenses [$]
0
2
4
6
8
10
12
Time


Expenses
[$]
25,000
0
Logistics
R&D
Marketing
0
2
4
6
8
10
12
Months since launch

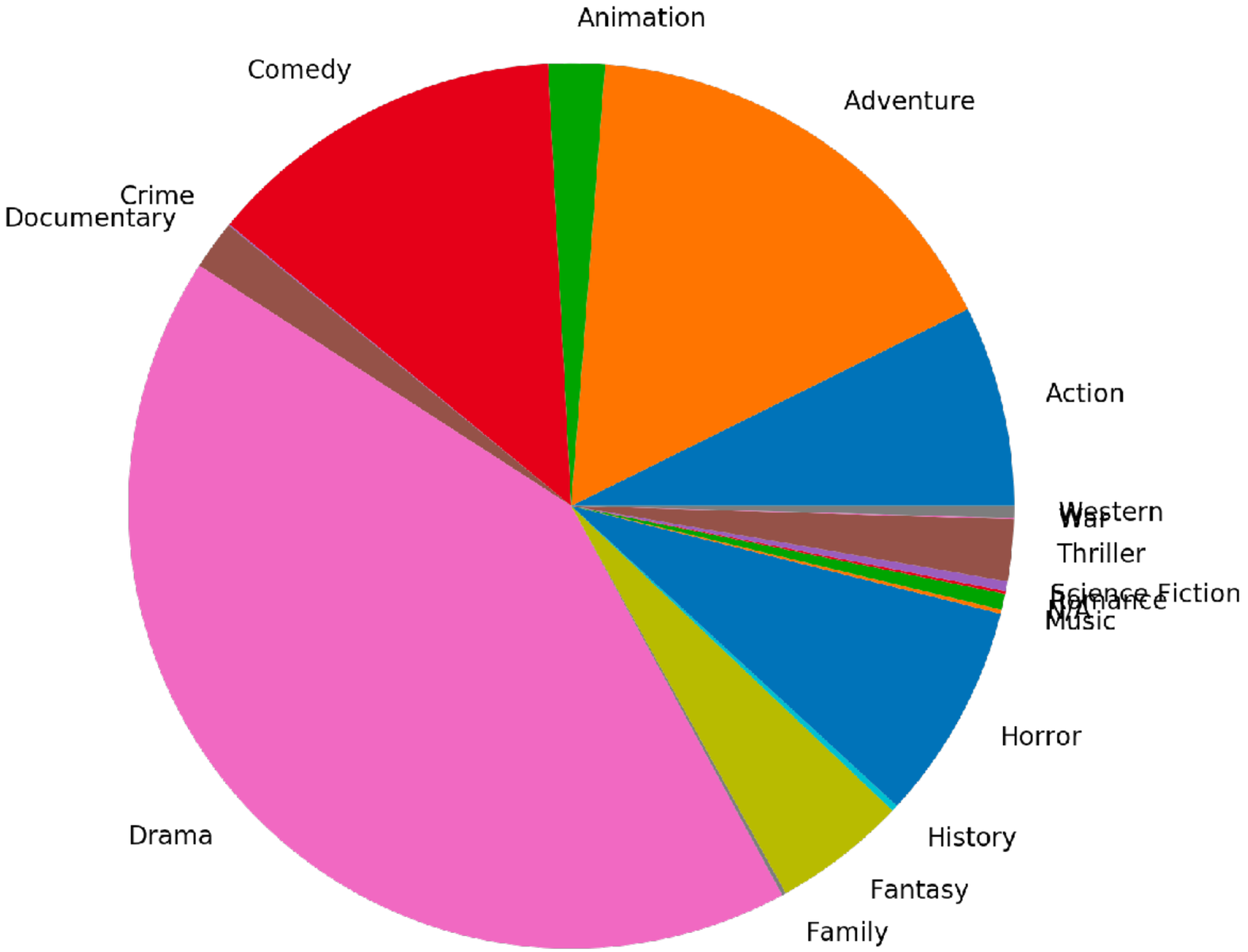

Expenses
[$]
25,000
0
0
2
4
6
8
10
12
Months since launch
Logistics: $25,757
R&D: $21,894
Marketing: $14,928

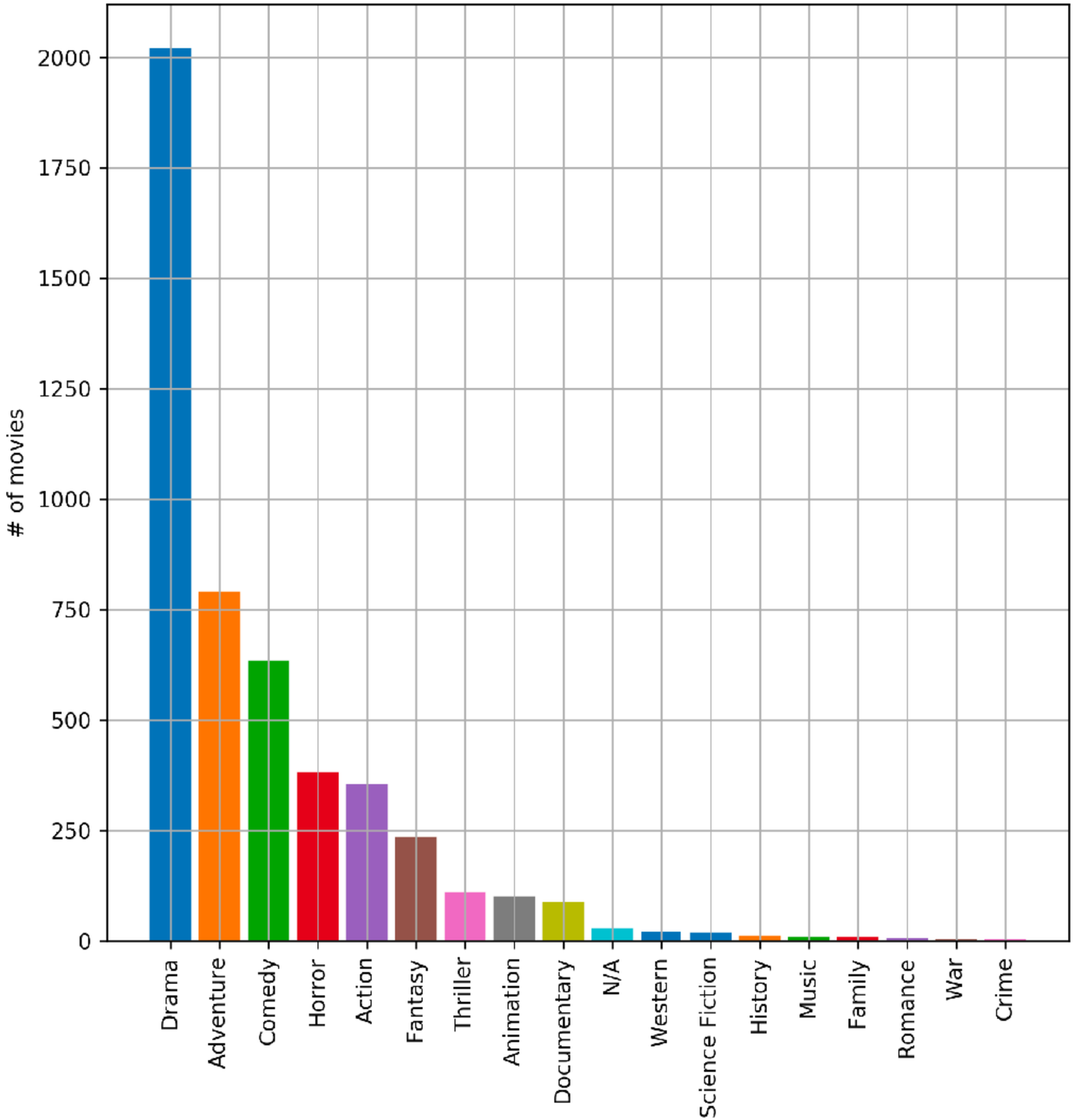

Expenses
[$]
25,000
0
0
2
4
6
8
10
12
Months since launch
Logistics: $25,757
R&D: $21,894
Marketing: $14,928

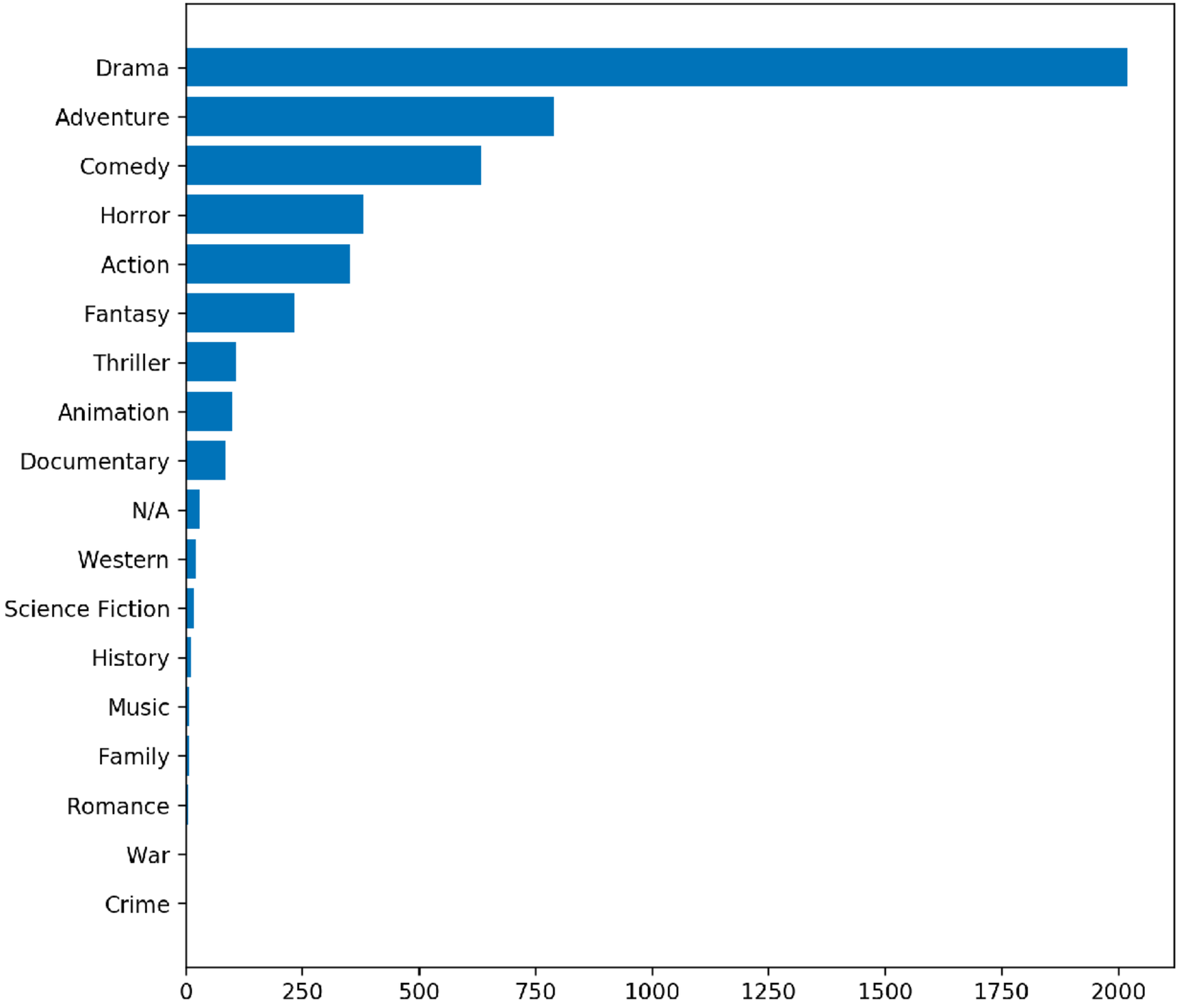
Three most common data visualization mistakes

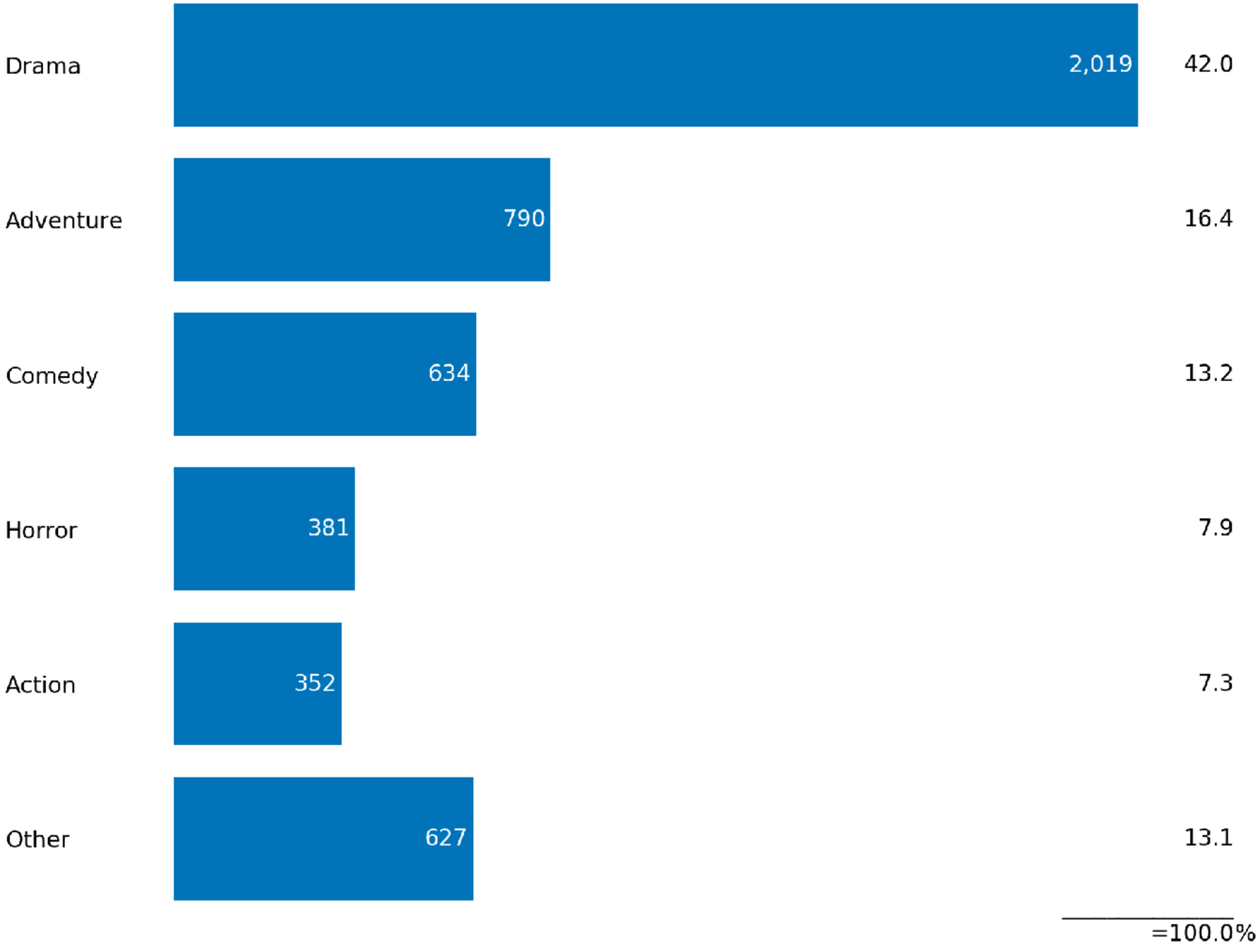
Graph contents and composition

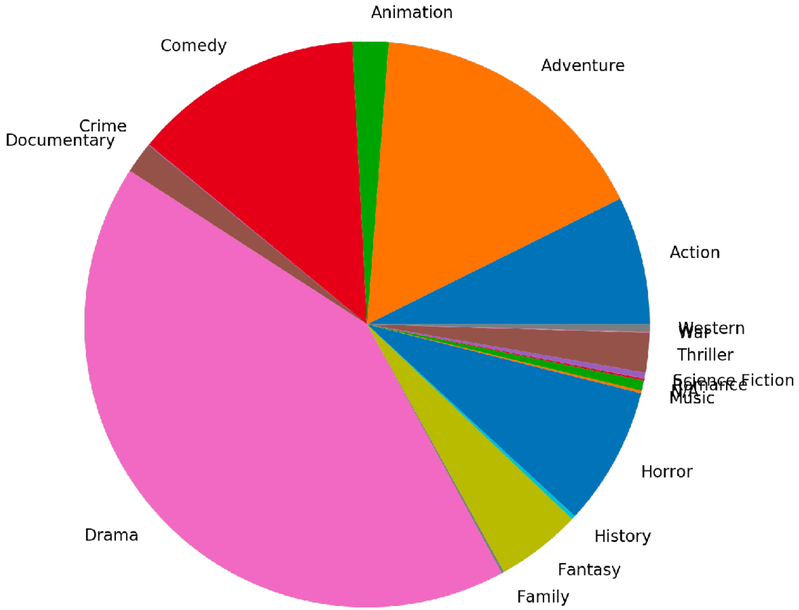
Optimize the data-ink ratio. Use appropriate graph types.

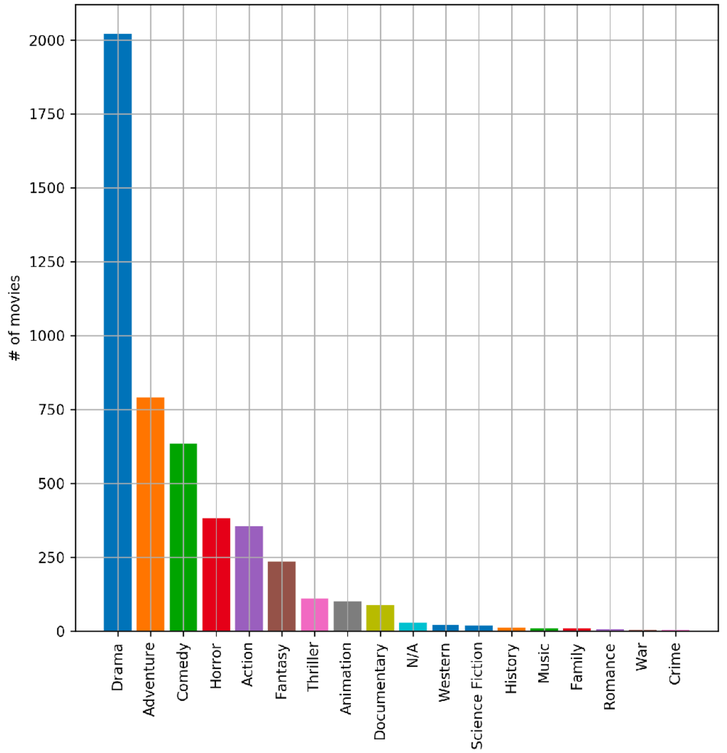

Animation
Comedy
Adventure
Crime
Documentary
Action
Western
War
Thriller
Science Fiction
Romance
N/A
Music
Horror
History
Fantasy
Family
Drama

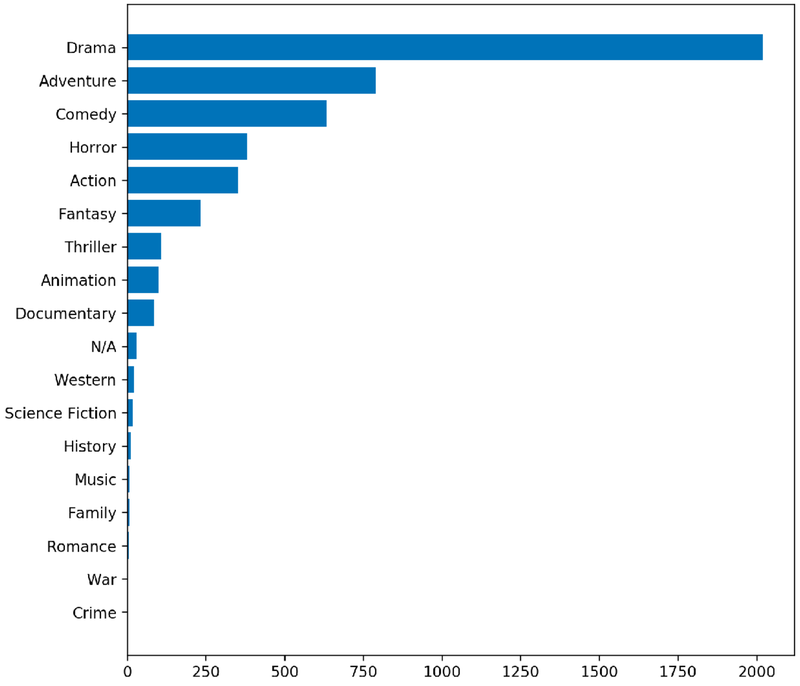

2000
1750
1500
1250
1000
750
500
250
0
# of movies
Drama
Adventure
Comedy
Horror
Action
Fantasy
Thriller
Animation
Documentary
N/A
Western
Science Fiction
History
Music
Family
Romance
War
Crime

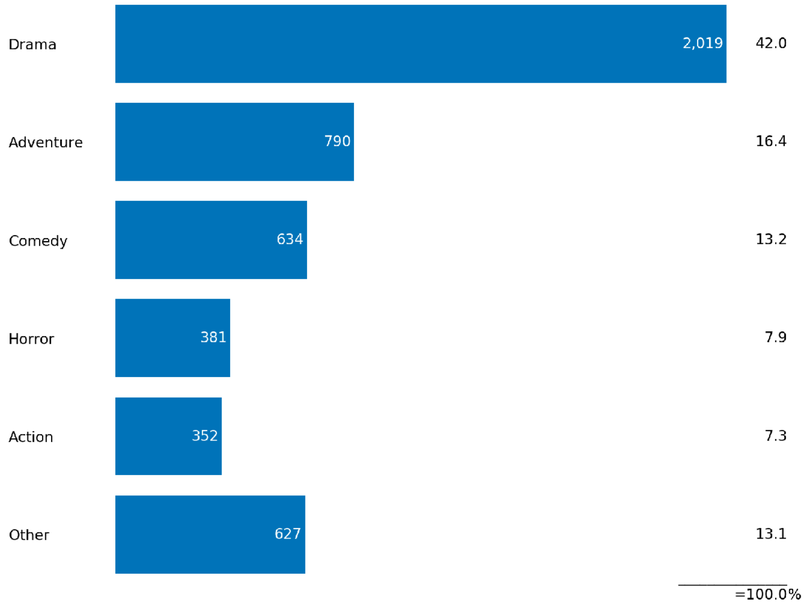

Drama
Adventure
Comedy
Horror
Action
Fantasy
Thriller
Animation
Documentary
N/A
Western
Science Fiction
History
Music
Family
Romance
War
Crime
0
250
500
750
1000
1250
1500
1750
2000


Drama
2,019
42.0
Adventure
790
16.4
Comedy
634
13.2
Horror
381
7.9
Action
352
7.3
Other
627
13.1
=100.0%


Animation
Comedy
Adventure
Crime
Documentary
Action
Western
War
Thriller
Science Fiction
Romance
N/A
Music
Horror
Drama
History
Fantasy
Family


# of movies
2000
1750
1500
1250
1000
750
500
250
0
Drama
Adventure
Comedy
Horror
Action
Fantasy
Thriller
Animation
Documentary
N/A
Western
Science Fiction
History
Music
Family
Romance
War
Crime


Drama
Adventure
Comedy
Horror
Action
Fantasy
Thriller
Animation
Documentary
N/A
Western
Science Fiction
History
Music
Family
Romance
War
Crime
0
250
500
750
1000
1250
1500
1750
2000


Drama
2,019
42.0
Adventure
790
16.4
Comedy
634
13.2
Horror
381
7.9
Action
352
7.3
Other
627
13.1
=100.0%

Three most common data visualization mistakes

Attitude

Evidence-based.Explanatory vs. Exploratory

Graph contents and composition

Optimize the data-ink ratio. Use appropriate graph types.

Three most common data visualization mistakes

3

RED GREEN BLUE RED BLUE GREEN

BLUE YELLOW RED RED BLUE GREEN

GREEN YELLOW RED GREEN BLUE

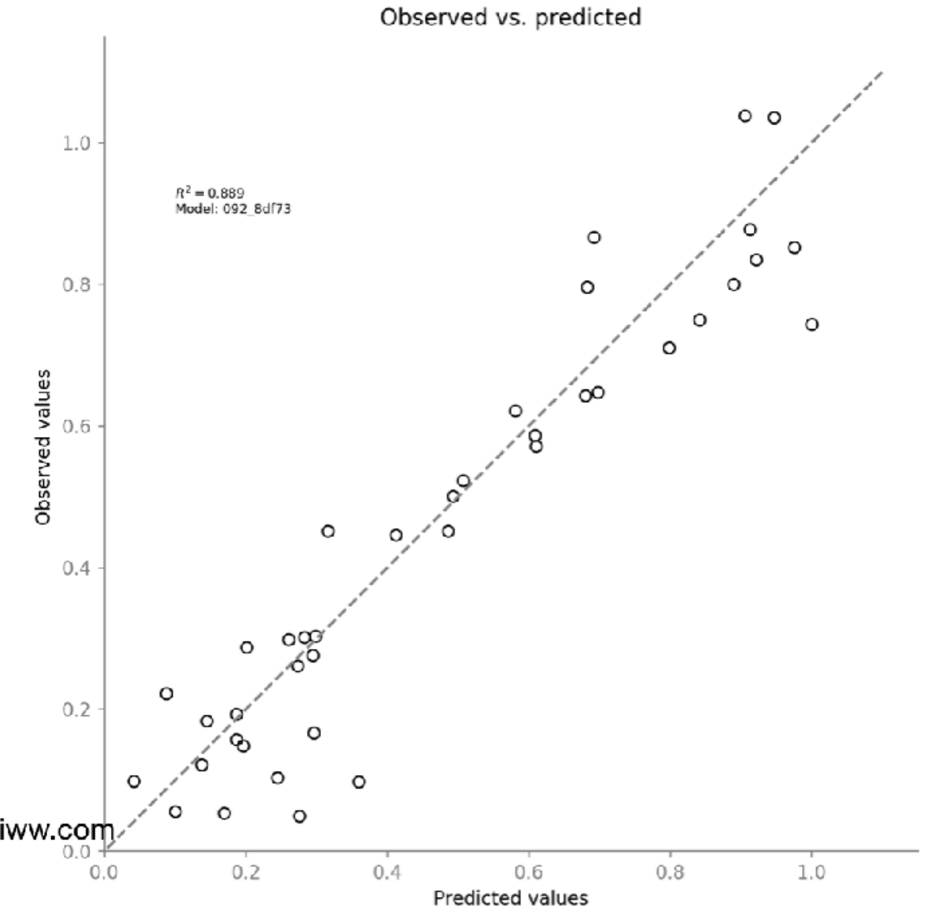
RED GREEN BLUE RED BLUE GREEN

BLUE YELLOW RED RED BLUE GREEN

GREEN YELLOW RED GREEN BLUE

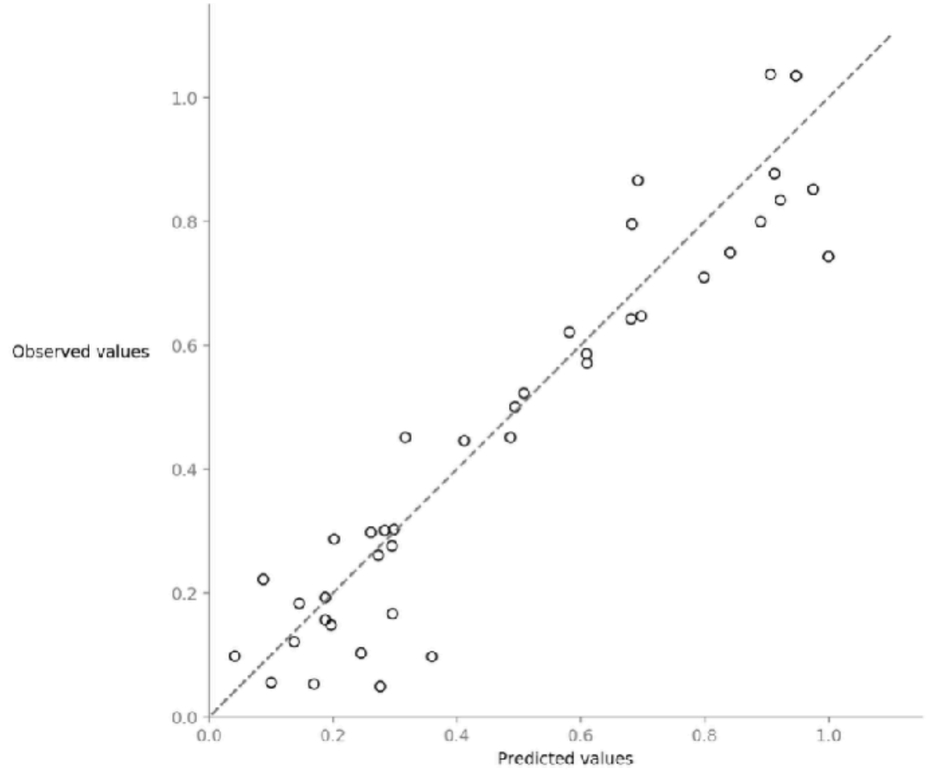
TEXT

Predicted vs. observed



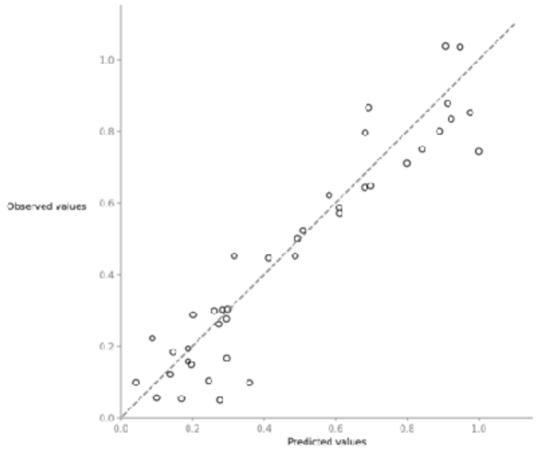
Data center: xylem.kdowiww.com
20170101-20171201

High agreement between the model and the observations

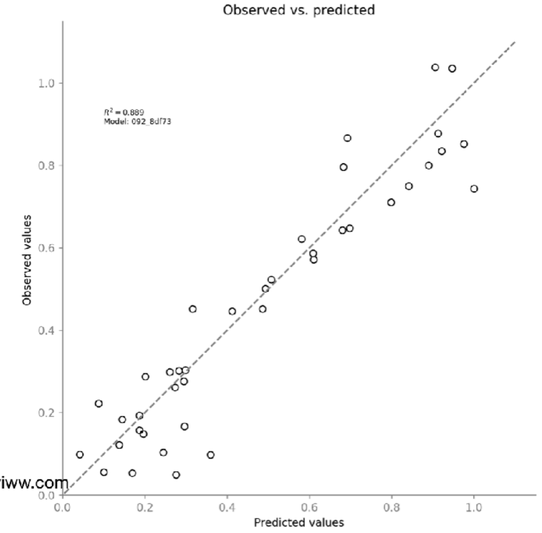


“So what?” vs. “What?”

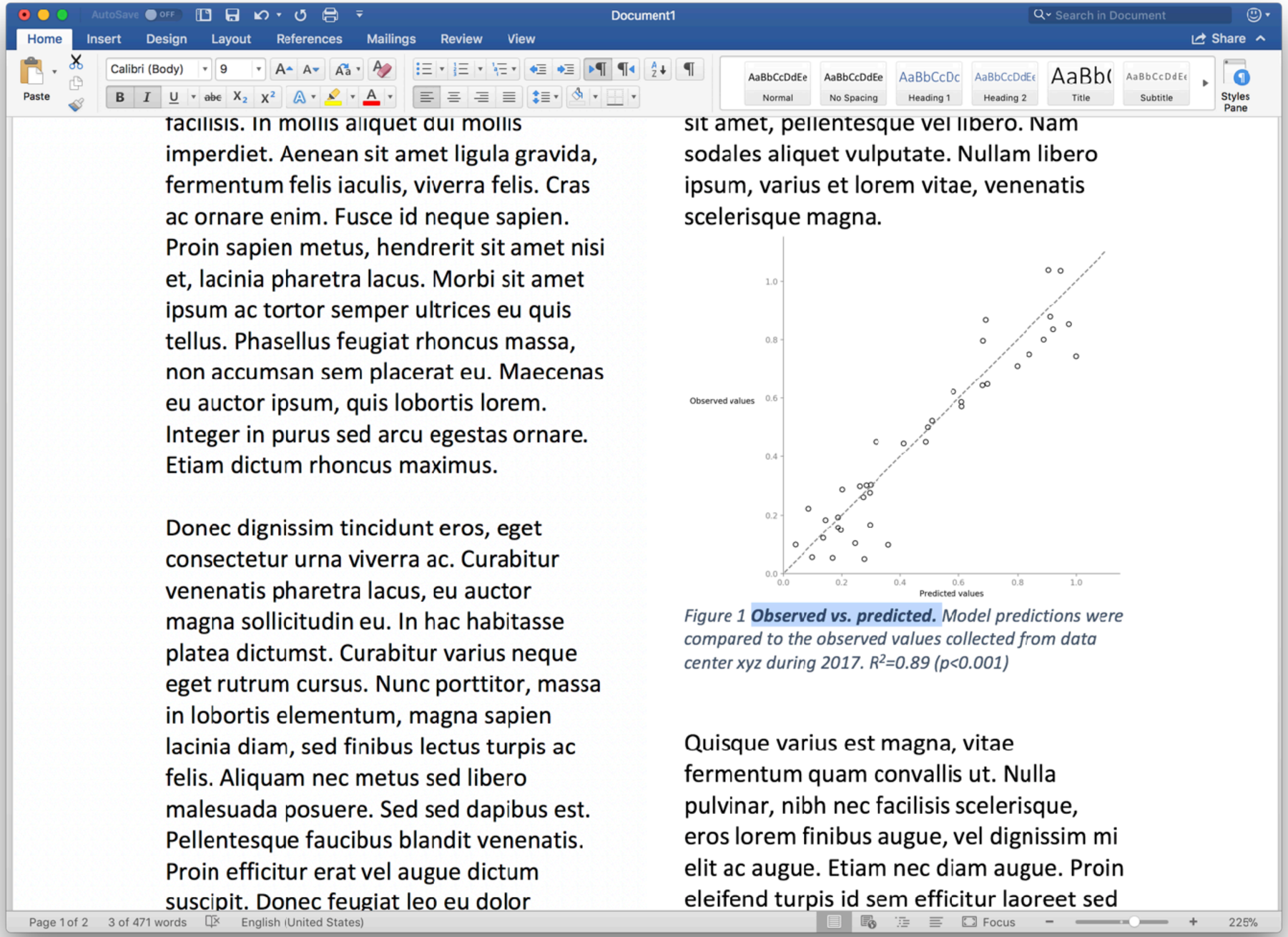
Good model performance
R^2=0.89



Predicted vs. observed



Data center: xylem.kdowiww.com
20170101-20171201


Document1
Search in Document
Home
Insert
Design
Layout
References
Mailings
Review
View
Share
Paste
Calibri (Body)
9
Normal
No Spacing
Heading 1
Heading 2
Title
Subtitle
Styles Pane
facilisis. In mollis aliquet dui mollis imperdiet. Aenean sit amet ligula gravida, fermentum felis iaculis, viverra felis. Cras ac ornare enim. Fusce id neque sapien. Proin sapien metus, hendrerit sit amet nisi et, lacinia pharetra lacus. Morbi sit amet ipsum ac tortor semper ultrices eu quis tellus. Phasellus feugiat rhoncus massa, non accumsan sem placerat eu. Maecenas eu auctor ipsum, quis lobortis lorem. Integer in purus sed arcu egestas ornare. Etiam dictum rhoncus maximus.
Donec dignissim tincidunt eros, eget consectetur urna viverra ac. Curabitur venenatis pharetra lacus, eu auctor magna sollicitudin eu. In hac habitasse platea dictumst. Curabitur varius neque eget rutrum cursus. Nunc porttitor, massa in lobortis elementum, magna sapien lacinia diam, sed finibus lectus turpis ac felis. Aliquam nec metus sed libero malesuada posuere. Sed sed dapibus est. Pellentesque faucibus blandit venenatis. Proin efficitur erat vel augue dictum suscipit. Donec feugiat leo eu dolor
sit amet, pellentesque vel libero. Nam sodales aliquet vulputate. Nullam libero ipsum, varius et lorem vitae, venenatis scelerisque magna.
Observed values
Predicted values
0.0
0.2
0.4
0.6
0.8
1.0
Figure 1 Observed vs. predicted. Model predictions were compared to the observed values collected from data center xyz during 2017. R2=0.89 (p<0.001)
Quisque varius est magna, vitae fermentum quam convallis ut. Nulla pulvinar, nibh nec facilisis scelerisque, eros lorem finibus augue, vel dignissim mi elit ac augue. Etiam nec diam augue. Proin eleifend turpis id sem efficitur laoreet sed
Page 1 of 2
3 of 471 words
English (United States)
Focus
225%

facilisis. In mollis aliquet dui mollis imperdiet. Aenean sit amet ligula gravida, fermentum felis iaculis, viverra felis. Cras ac ornare enim. Fusce id neque sapien. Proin sapien metus, hendrerit sit amet nisi et, lacinia pharetra lacus. Morbi sit amet ipsum ac tortor semper ultrices eu quis tellus. Phasellus feugiat rhoncus massa, non accumsan sem placerat eu. Maecenas eu auctor ipsum, quis lobortis lorem. Integer in purus sed arcu egestas ornare. Etiam dictum rhoncus maximus.

Donec dignissim tincidunt eros, eget consectetur urna viverra ac. Curabitur venenatis pharetra lacus, eu auctor magna sollicitudin eu. In hac habitasse platea dictumst. Curabitur varius neque eget rutrum cursus. Nunc porttitor, massa in lobortis elementum, magna sapien lacinia diam, sed finibus lectus turpis ac felis. Aliquam nec metus sed libero malesuada posuere. Sed sed dapibus est. Pellentesque faucibus blandit venenatis. Proin efficitur erat vel augue dictum suscipit. Donec feugiat leo eu dolor

sit amet, pellentesque vel libero. Nam sodales aliquet vulputate. Nullam libero ipsum, varius et lorem vitae, venenatis scelerisque magna.

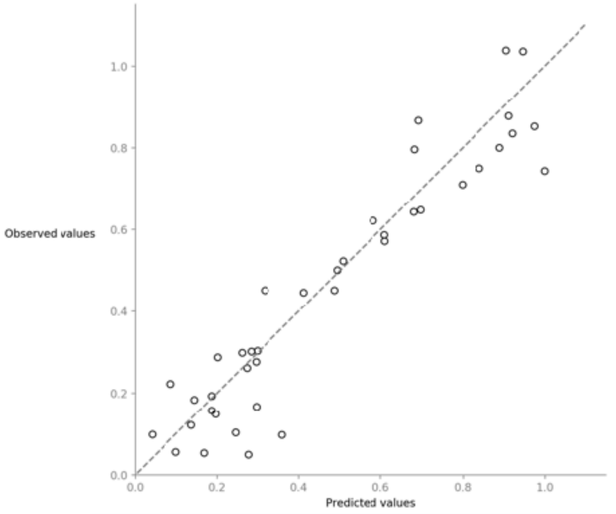
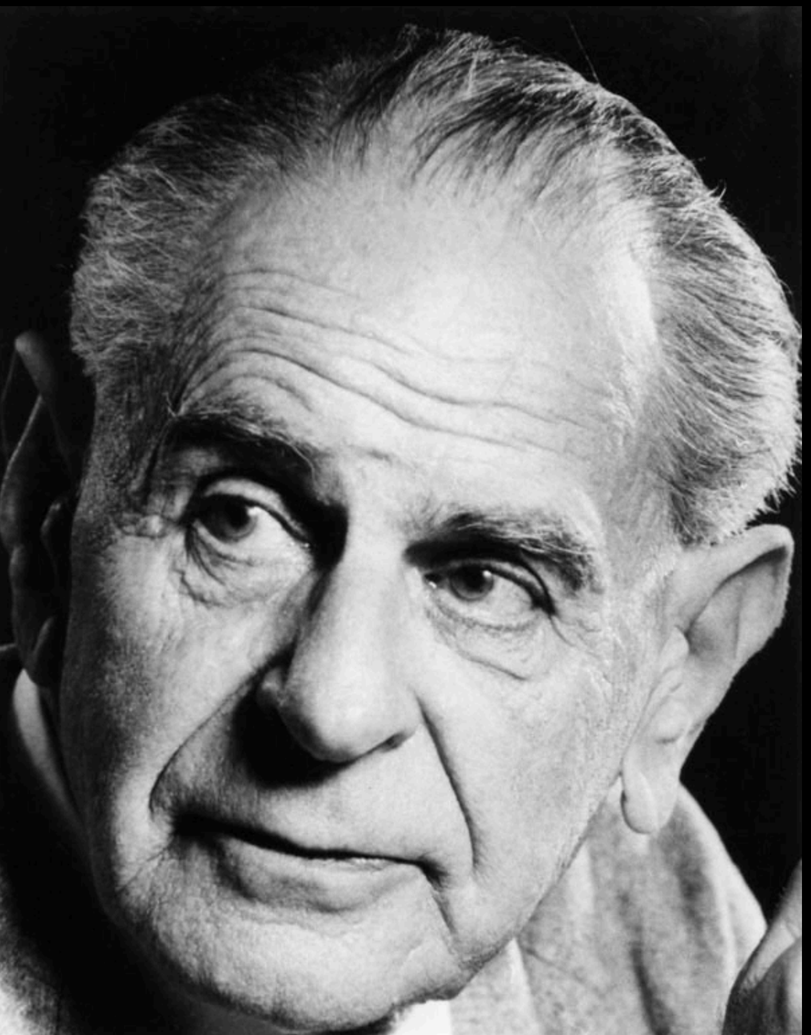


Figure 1 ***Good agreement between the model and the observations.*** *R^2=0.89 ($p<0.001$) Model predictions were compared to the observed values collected from data center xyz during 2017.*

Quisque varius est magna, vitae fermentum quam convallis ut. Nulla pulvinar, nibh nec facilisis scelerisque, eros lorem finibus augue, vel dignissim mi elit ac augue. Etiam nec diam augue. Proin



Karl Popper
1902-1994

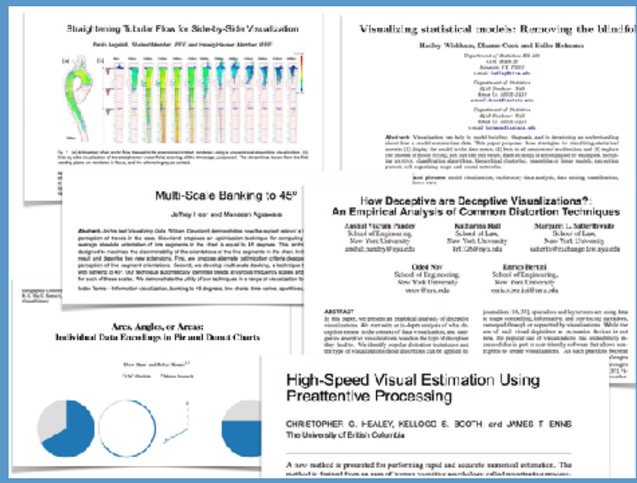
Three most common data visualization mistakes

Three most common data visualization mistakes

Attitude

Evidence-based

Explanatory vs. Exploratory

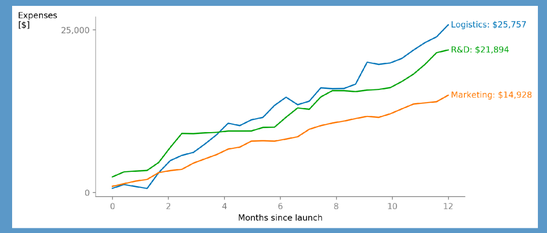


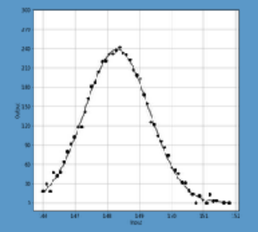
Three most common data visualization mistakes

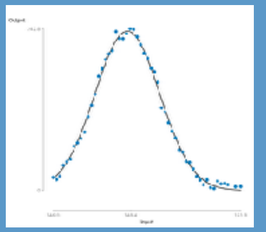
Graph contents and composition

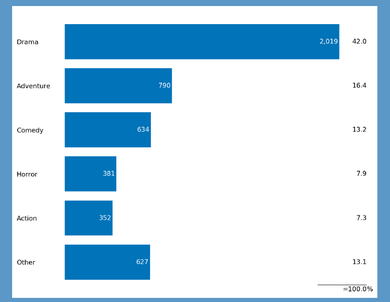
Remove, remove, remove.

Use appropriate graph types.





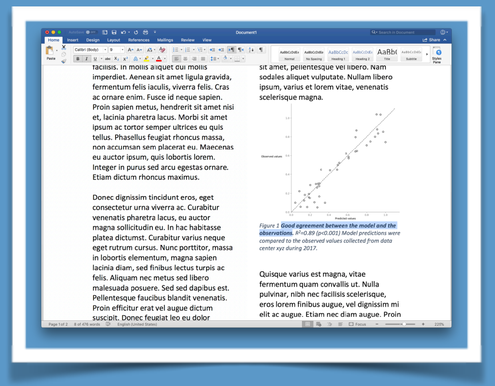




Three most common data visualization mistakes

Conclusions matter

”So what?” is more important than “what?”



Data visualization as an engineering task

a methodological approach towards creating effective data visualization

Boris Gorelik

http://gorelik.net

AUTOMATTIC

WordPress.com

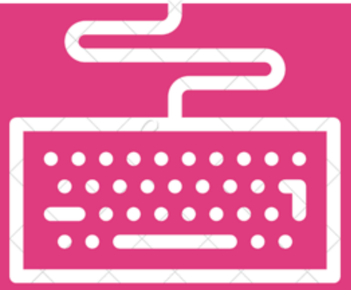
AZRIELI College of Engineering Jerusalem

"First, go straight. Turn right after half a mile, and continue walking until you see it"
— Anonymous

Boris Gorelik, Ph.D.

The ABCD of data visualization

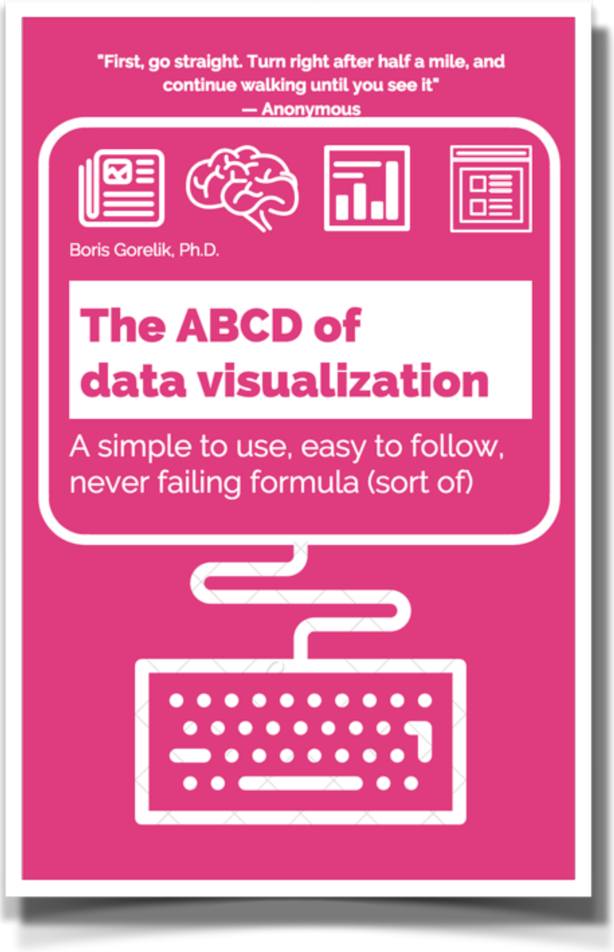
A simple to use, easy to follow, never failing formula (sort of)



Data visualization as an engineering task

Audience

who is your target audience?



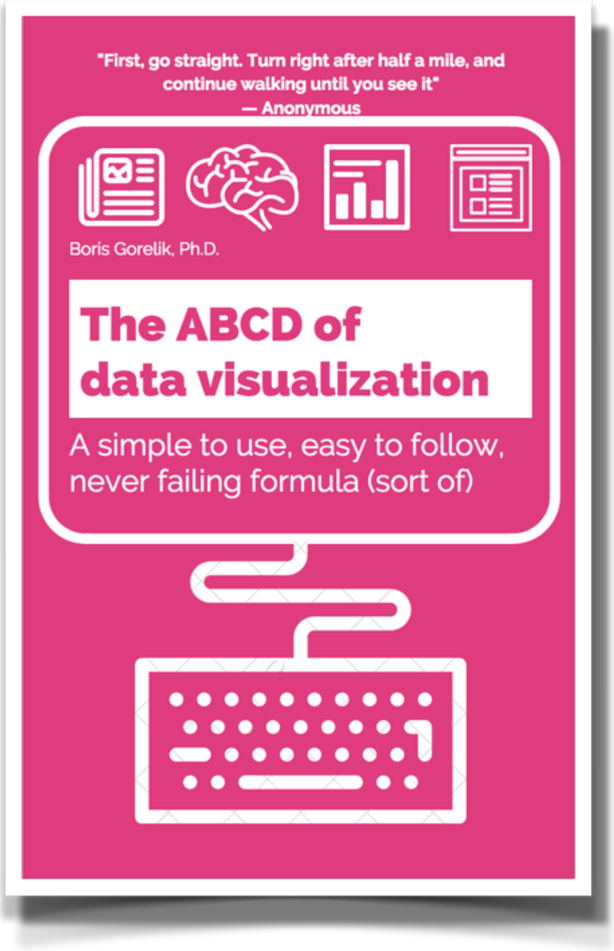
Data visualization as an engineering task

Audience

who is your target audience

Build the graph

using the right data and the right graph type



Data visualization as an engineering task

Audience

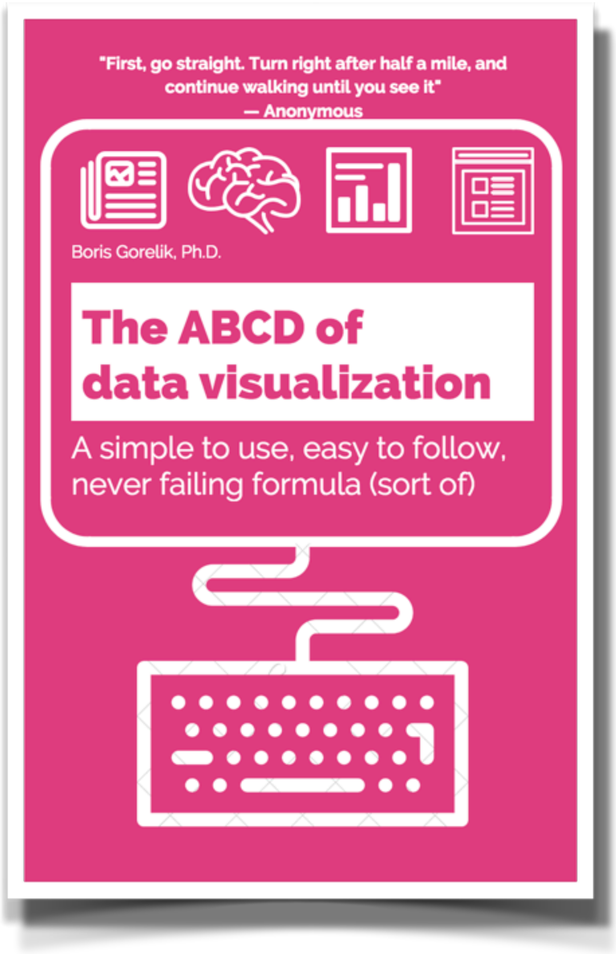
who is your target audience

Build the graph

using the right data and the right graph type

Conclusion

does the graph say what it says that it says?



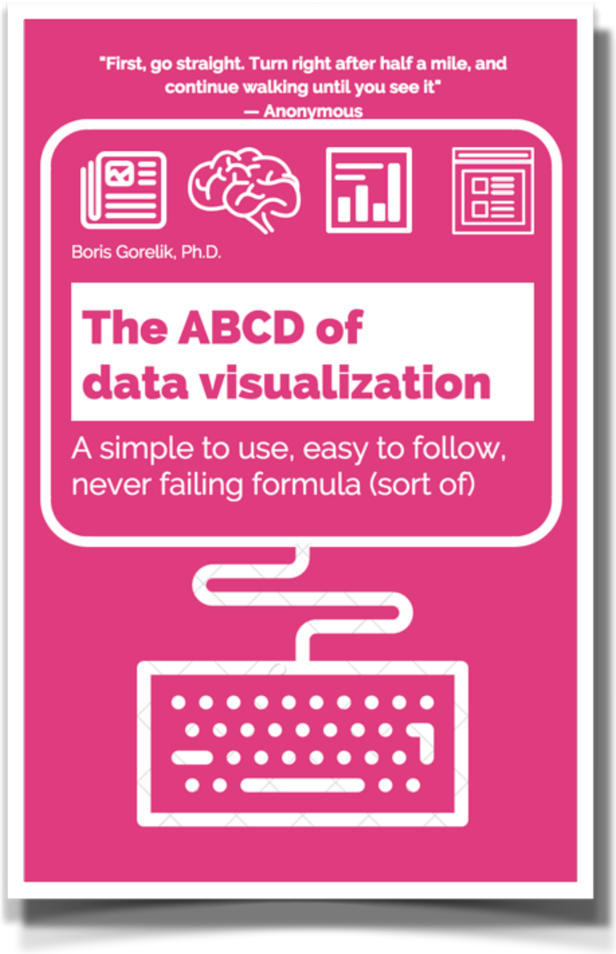
Data visualization as an engineering task

Audience
who is your target audience

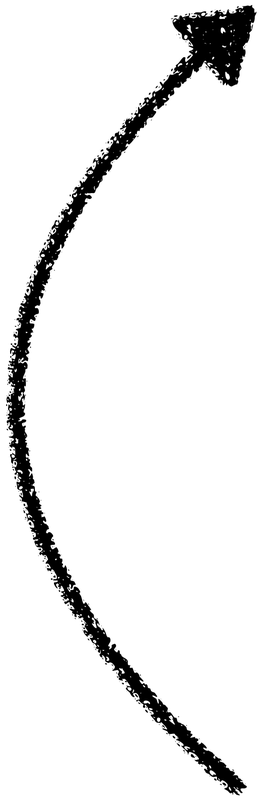
Build the graph
using the right data and the right graph type

Conclusion
does the graph say what it says that it says?

Delete
remove everything unneeded



Data visualization as an engineering task



Audience

who is your target audience

Build the graph

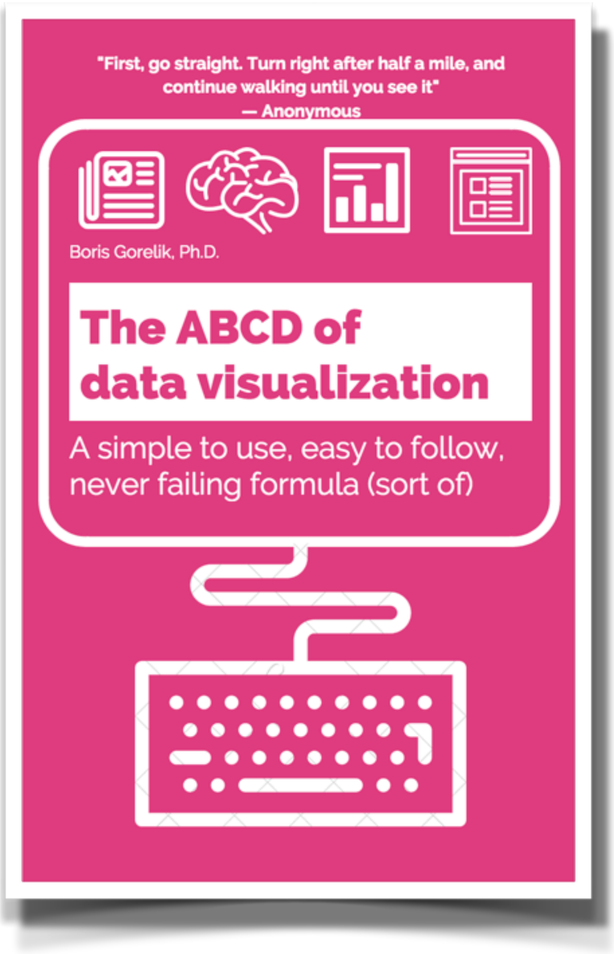
using the right data and the right graph type

Conclusion

does the graph say what it says that it says?

Delete

remove everything unneeded

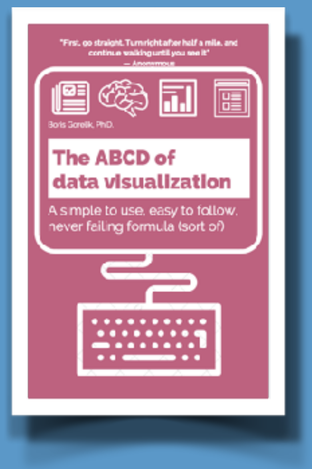


Data visualization as an engineering task

Attitude
Evidence-based.Explanatory vs. Exploratory

Graph contents and composition
Remove, remove, remove. Use appropriate graph types.

Conclusions matter
”So what?” is more important than “what?”



Audience
who is your target audience

Build the graph
using the right data and the right graph type

Conclusion
does the graph say what it says that it says?

Delete
remove everything unneeded

Boris Gorelik, Ph.D.
boris@gorelik.net
http://gorelik.net

AUTOMATTIC
WordPress.com

Reading material