Data Visualisation:

Telling Visual Stories From Data

Bevan Koopman, PhD
Research Scientist

Outline

- What is data visualisation and why it matters?
- The grammar of graphical elements
- Choosing appropriate visual cues
- Structuring data and workflows that support data vis
- Practical guide to plot types

About me

2014-2017: Research Scientist

2010-2013: PhD, Information Retrieval

2005-2009: Software Engineer

2002-2005: Research Engineer

1999-2002: BInfoTech (Hons)

What do I work on?

- Information retrieval - the science behind search engines
  - Semantic search - breaking the dependence on terms
- Health informatics - high impact search
  - How clinician search

THE AUSTRALIAN E-HEALTH RESEARCH CENTRE

Data Vis. in my work

Data Vis. in my work

What Makes an Effective Clinical Query and Querier?

Bevan Koopman
Australian e-Health Research Centre, CSIRO, Brisbane, Australia

Guido Zuccon
School of Electrical Engineering & Computer Science, Queensland University of Technology, Brisbane, Australia

Peter Bruza
School of Information Systems, Queensland University of Technology, Brisbane, Australia

In this paper, we perform an in-depth study into how clinicians represent their information needs and the influence this has on information retrieval (IR) effectiveness. While considerable research in IR has considered the effectiveness of IR systems, there is still a considerable gap in the understanding of how users contribute to the effectiveness of these systems. The paper aims to contribute to this by studying how clinicians search for information.
Multiple representations of a information need — from verbose patient case descriptions to ad-hoc queries — were considered in order to understand their effect on retrieval. Four clinicians provided queries and performed relevance assessment to form a test collection used in this study. The different query formulation strategies of each clinician, and their effectiveness, were investigated.
The results show that query formulation had more impact on retrieval effectiveness than the particular retrieval systems used. The most effective queries were short, ad-hoc keyword queries. Different clinicians were observed to consistently adopt specific query formulation strategies. The most effective queriers were those who, given their information need, inferred novel keywords most likely to appear in relevant documents.
This study reveals aspects of how people search within the clinical domain. This can help inform the development of new models and methods that specifically focus on the query formulation process to improve retrieval effectiveness.

Introduction

Information retrieval (IR) systems have the fundamental purpose of promoting interactions with information that support people to achieve their goals and agendas in a wide variety of situations. Although it has long been held that IR research and practice must be based on an understanding of the people for whom the systems are intended, there is still a large gap between the study of the users of the systems as opposed to the algorithms underpinning the systems and how these are evaluated (James Allan and (eds.), 2012). This paper aims to be a stepping stone in bridging this gap by understanding how clinicians engage in medical information retrieval. In particular, our concern is both to understand what makes a good clinical *query* as well as to understand what makes a good clinical *querier*.

In this paper, we perform an in-depth study into how clinicians represent their information needs and the influence this had on retrieval effectiveness. Unlike the standard approach in IR evaluation of a single query per information need, we considered three different representations of an information need: i) verbose patient case descriptions (78 words per topic); ii) shorter patient case summaries (22 words per topic) of the patient case description; and iii) short ad-hoc queries (4.2 words per topic) expressed by clinicians. All three representations were realistic queries taken from a real-word clinical search scenario. These multiple representations of a single information need were used to retrieve clinical documents via a number of retrieval systems. Medical professionals were employed to provide relevance assessments of the retrieved documents, which allows the effectiveness of different query representations to be studied. In addition, we studied the different query formulation strategies of different clinicians to understand what constituted an effective clinical querier. More specifically, this paper aims to answer the following overarching questions:

What makes a good clinical query?

- How did different representations of the same information need — from ad-hoc queries through to verbose patient case descriptions — influence retrieval effectiveness? Were human-derived ad-hoc queries more effective than verbose case descriptions?
- What was the variation in effectiveness for different ad-hoc queries and for different clinicians? When was ad-hoc querying best or worst?

What makes a good clinical querier?

- To what extend did clinicians select keywords from the patient case description to form their ad-hoc queries; or did they derive unique keywords? Which method was more effective?
- Are there specific query strategies between clinicians that proved more effective?

Our findings confirm that different representations of information needs did indeed have a large impact on retrieval effectiveness. We find that humans were capable of formulating very effective queries but that there was large differences in effectiveness between

Data Vis. in my work

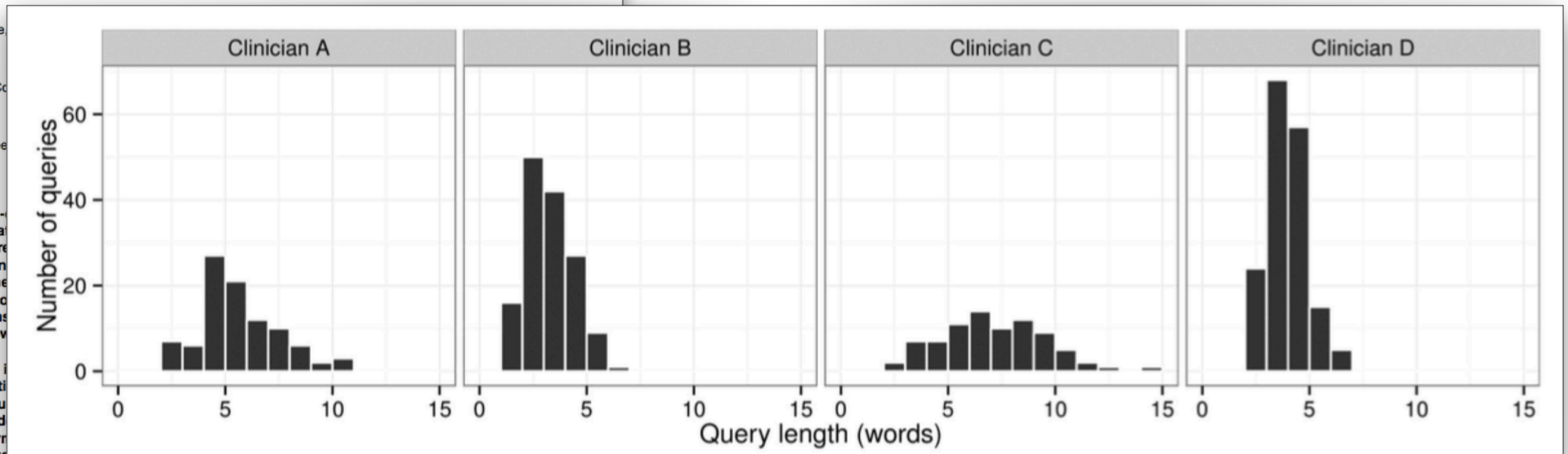
What Makes an Effective Clinical Query and Querier?

Bevan Koopman
Australian e-Health Research Centre,

Guido Zuccon
School of Electrical Engineering & Co

Peter Bruza
School of Information Systems, Quee

**In this paper, we perform an in-
icians represent their informa
ence this has on information re
While considerable research in
fectiveness of IR systems, the
gap in the understanding of ho
effectiveness of these systems
tribute to this by studying how
formation.
Multiple representations of a i
verbose patient case descripti
were considered in order to u
retrieval. Four clinicians provid
relevance assessment to forr
in this study. The different que
of each clinician, and their effectiveness, were investigated.
The results show that query formulation had more impact on retrieval effectiveness than the particular retrieval systems used. The most effective queries were short, ad-hoc keyword queries. Different clinicians were observed to consistently adopt specific query formulation strategies. The most effective queriers were those who, given their information need, inferred novel keywords most likely to appear in relevant documents.
This study reveals aspects of how people search within the clinical domain. This can help inform the development of new models and methods that specifically focus on the query formulation process to improve retrieval effectiveness.**



Introduction

Information retrieval (IR) systems have the fundamental purpose of promoting interactions with information that support people to achieve their goals and agendas in a wide variety of situations. Although it has long been held that IR research and practice must be based on an understanding of the people for whom the systems are intended, there is still a large gap between the study of the users of the systems as opposed to the algorithms underpinning the systems and how these are evaluated (James Allan and (eds.), 2012). This paper aims to be a stepping stone in bridging this gap by understanding how clinicians engage in medical information retrieval. In were employed to provide relevance assessments of the retrieved documents, which allows the effectiveness of different query representations to be studied. In addition, we studied the different query formulation strategies of different clinicians to understand what constituted an effective clinical querier. More specifically, this paper aims to answer the following overarching questions:

What makes a good clinical query?

- How did different representations of the same information need — from ad-hoc queries through to verbose patient case descriptions — influence retrieval effectiveness? Were human-derived ad-hoc queries more effective than verbose case descriptions?
- What was the variation in effectiveness for different ad-hoc queries and for different clinicians? When was ad-hoc querying best or worst?

What makes a good clinical querier?

- To what extend did clinicians select keywords from the patient case description to form their ad-hoc queries; or did they derive unique keywords? Which method was more effective?
- Are there specific query strategies between clinicians that proved more effective?

Our findings confirm that different representations of information needs did indeed have a large impact on retrieval effectiveness. We find that humans were capable of formulating very effective queries but that there was large differences in effectiveness between

Data Vis. in my work

What Makes an Effective Clinical Query and Querier?

Bevan Koopman
Australian e-Health Research Centre,

Guido Zuccon
School of Electrical Engineering & Co

Peter Bruza
School of Information Systems, Quee

**In this paper, we perform an in-
icians represent their informa
ence this has on information re
While considerable research in
fectiveness of IR systems, the
gap in the understanding of ho
effectiveness of these systems
tribute to this by studying how
formation.
Multiple representations of a i
verbose patient case descripti
were considered in order to u
retrieval. Four clinicians provid
relevance assessment to forr
in this study. The different que
of each clinician, and their effectiveness, were investigated.
The results show that query formulation had more impact on retrieval effectiveness than the particular retrieval systems used. The most effective queries were short, ad-hoc keyword queries. Different clinicians were observed to consistently adopt specific query formulation strategies. The most effective queriers were those who, given their information need, inferred novel keywords most likely to appear in relevant documents.
This study reveals aspects of how people search within the clinical domain. This can help inform the development of new models and methods that specifically focus on the query formulation process to improve retrieval effectiveness.**

Introduction

Information retrieval (IR) systems have the fundamental purpose of promoting interactions with information that support people to achieve their goals and agendas in a wide variety of situations. Although it has long been held that IR research and practice must be based on an understanding of the people for whom the systems are intended, there is still a large gap between the study of the users of the systems as opposed to the algorithms underpinning the systems and how these are evaluated (James Allan and (eds.), 2012). This paper aims to be a stepping stone in bridging this gap by understanding how clinicians engage in medical information retrieval. In

were employed to provide relevance assessments of the retrieved documents, which allows the effectiveness of different query representations to be studied. In addition, we studied the different query formulation strategies of different clinicians to understand what constituted an effective clinical querier. More specifically, this paper aims to answer the following overarching questions:

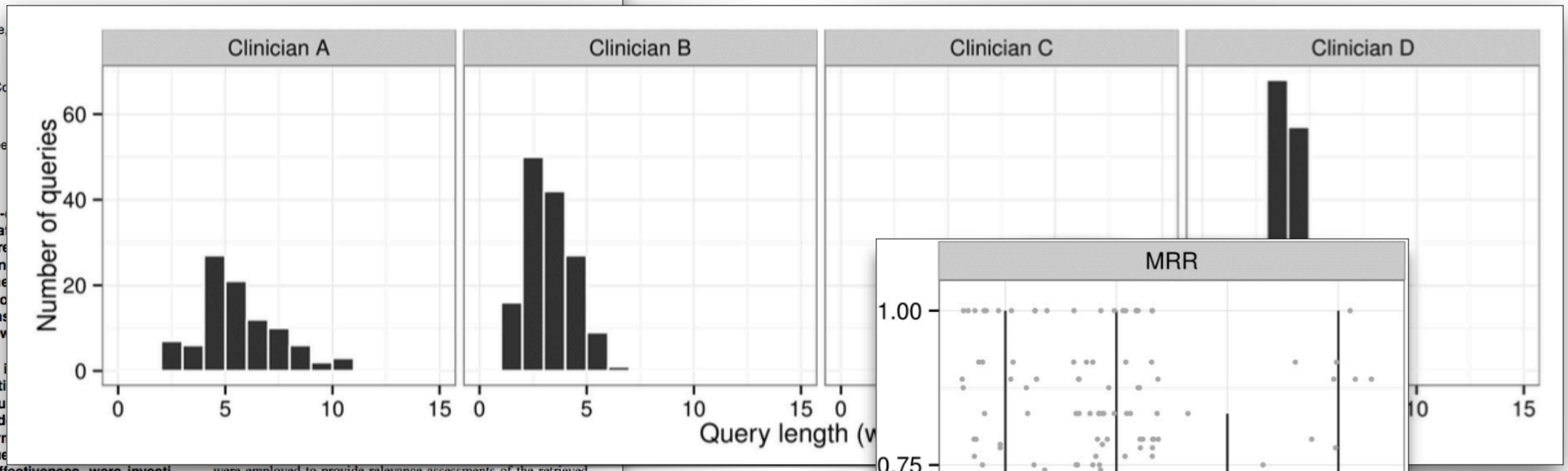
What makes a good clinical query?

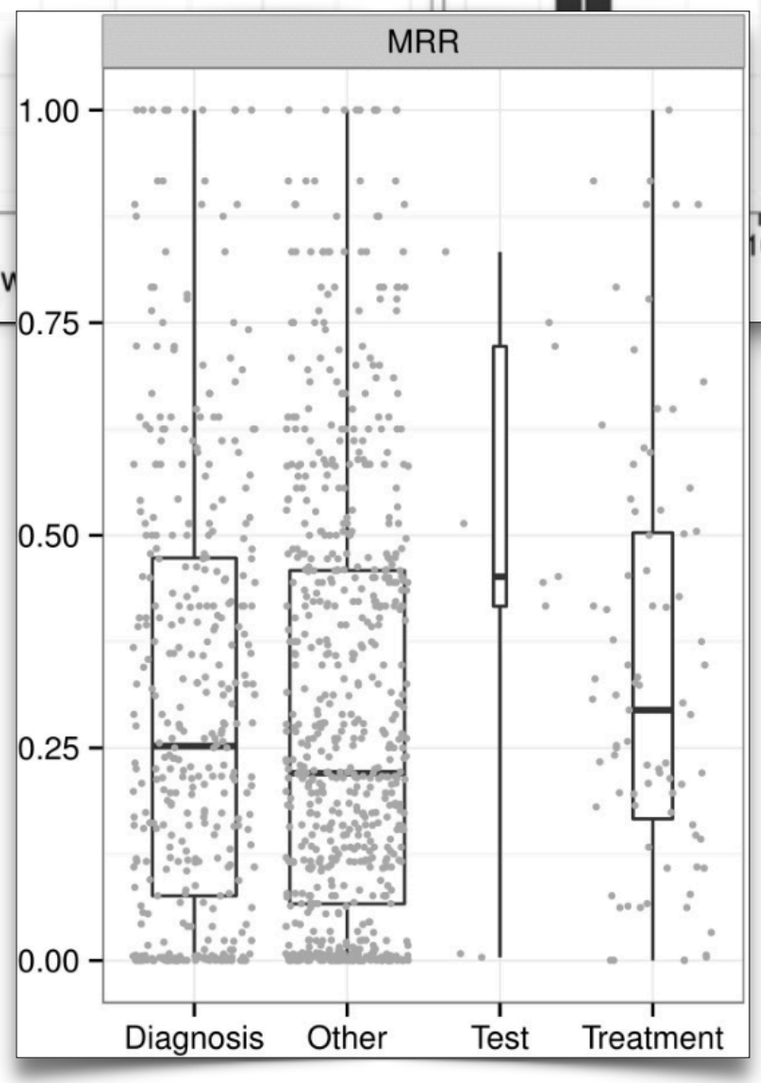
- How did different representations of the same information need — from ad-hoc queries through to verbose patient case descriptions — influence retrieval effectiveness? Were human-derived ad-hoc queries more effective than verbose case descriptions?
- What was the variation in effectiveness for different ad-hoc queries and for different clinicians? When was ad-hoc querying best or worst?

What makes a good clinical querier?

- To what extend did clinicians select keywords from the patient case description to form their ad-hoc queries; or did they derive unique keywords? Which method was more effective?
- Are there specific query strategies between clinicians that proved more effective?

Our findings confirm that different representations of information needs did indeed have a large impact on retrieval effectiveness. We find that humans were capable of formulating very effective queries but that there was large differences in effectiveness between





Data Vis. in my work

Inf Retrieval J (2016) 19:6–37
DOI 10.1007/s10791-015-9268-9

MEDICAL INFORMATION RETRIEVAL

Information retrieval as semantic inference: a Graph Inference model applied to medical search

Bevan Koopman[1] · Guido Zuccon[2] · Peter Bruza[2] · Laurianne Sitbon[2] · Michael Lawley[1]

Received: 15 December 2014 / Accepted: 7 September 2015 / Published online: 20 November 2015
© Springer Science+Business Media New York 2015

Abstract This paper presents a Graph Inference retrieval model that integrates structured knowledge resources, statistical information retrieval methods and inference in a unified framework. Key components of the model are a graph-based representation of the corpus and retrieval driven by an *inference* mechanism achieved as a traversal over the graph. The model is proposed to tackle the semantic gap problem—the mismatch between the raw data and the way a human being interprets it. We break down the semantic gap problem into five core issues, each requiring a specific type of inference in order to be overcome. Our model and evaluation is applied to the medical domain because search within this domain is particularly challenging and, as we show, often requires inference. In addition, this domain features both structured knowledge resources as well as unstructured text. Our evaluation shows that inference can be effective, retrieving many new relevant documents that are not retrieved by state-of-the-art information retrieval models. We show that many retrieved documents were not pooled by keyword-based search methods, prompting us to perform additional relevance assessment on these new documents. A third of the newly retrieved documents judged were found to be relevant. Our analysis provides a thorough understanding of when and how to apply inference for retrieval, including a categorisation of queries according to the effect of inference. The inference mechanism promoted recall

Data Vis.

Inf Retrieval J (2016) 19:6–37
DOI 10.1007/s10791-015-9268-9

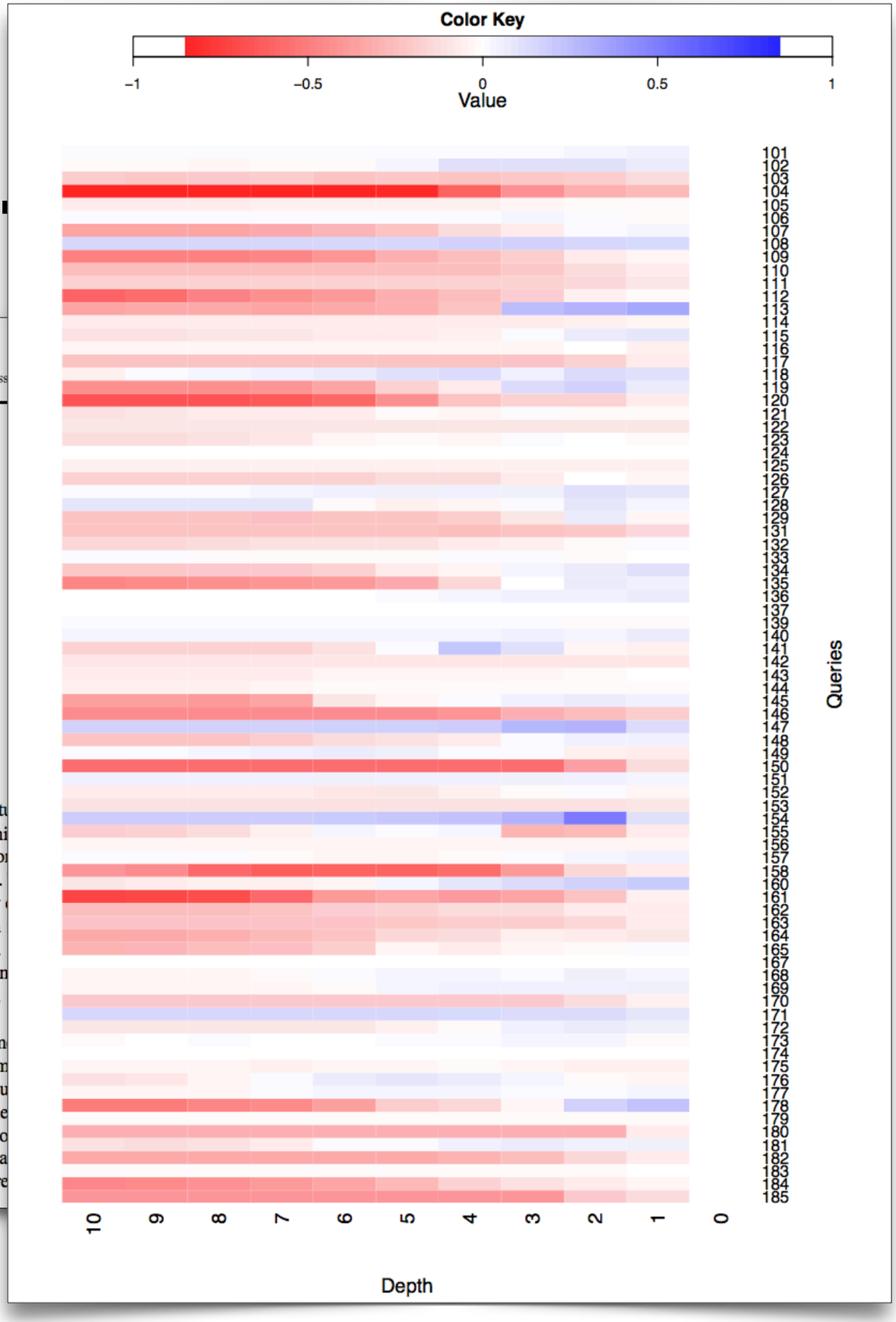
MEDICAL INFORMATION RETRIEVAL

Information retrieval as semantic inference: a Graph Inference model applied to medical search

Bevan Koopman[1] · Guido Zuccon[2] · Peter Bruza[2] · Laurianne Sitbon[2] · Michael Lawley[1]

Received: 15 December 2014 / Accepted: 7 September 2015 / Published online: 20 November 2015
© Springer Science+Business Media New York 2015

Abstract This paper presents a Graph Inference retrieval model that integrates structu
knowledge resources, statistical information retrieval methods and inference in a uni
framework. Key components of the model are a graph-based representation of the co
and retrieval driven by an *inference* mechanism achieved as a traversal over the graph.
model is proposed to tackle the semantic gap problem—the mismatch between the raw
and the way a human being interprets it. We break down the semantic gap problem
five core issues, each requiring a specific type of inference in order to be overcome.
model and evaluation is applied to the medical domain because search within this do
is particularly challenging and, as we show, often requires inference. In addition,
domain features both structured knowledge resources as well as unstructured text.
evaluation shows that inference can be effective, retrieving many new relevant docum
that are not retrieved by state-of-the-art information retrieval models. We show that m
retrieved documents were not pooled by keyword-based search methods, prompting u
perform additional relevance assessment on these new documents. A third of the ne
retrieved documents judged were found to be relevant. Our analysis provides a thoro
understanding of when and how to apply inference for retrieval, including a categorisa
of queries according to the effect of inference. The inference mechanism promoted re



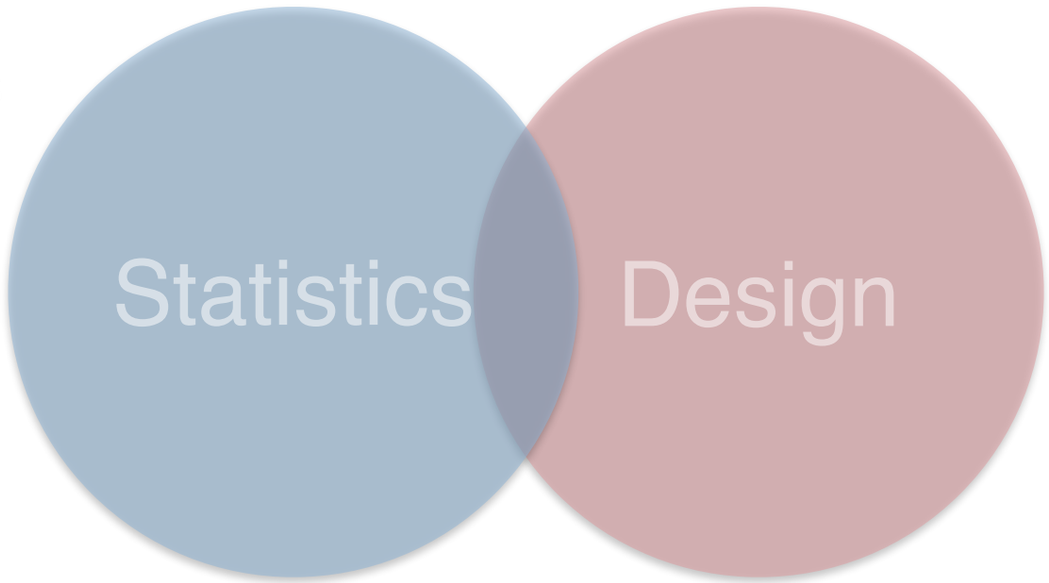
What is data visualisation?

- Well, an important skill set for any Data Scientist
- A combination…

What is data visualisation?

- Well, an important skill set for any Data Scientist
- A combination…

Graphics data analysis

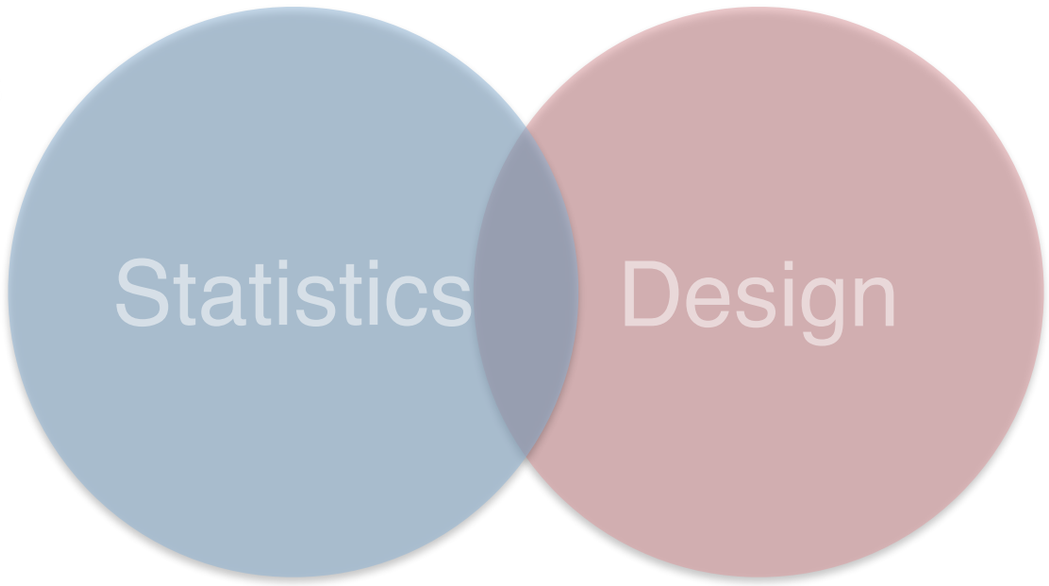


Communication & perception of results

What is data visualisation?

- Well, an important skill set for any Data Scientist
- A combination…

Graphics data analysis

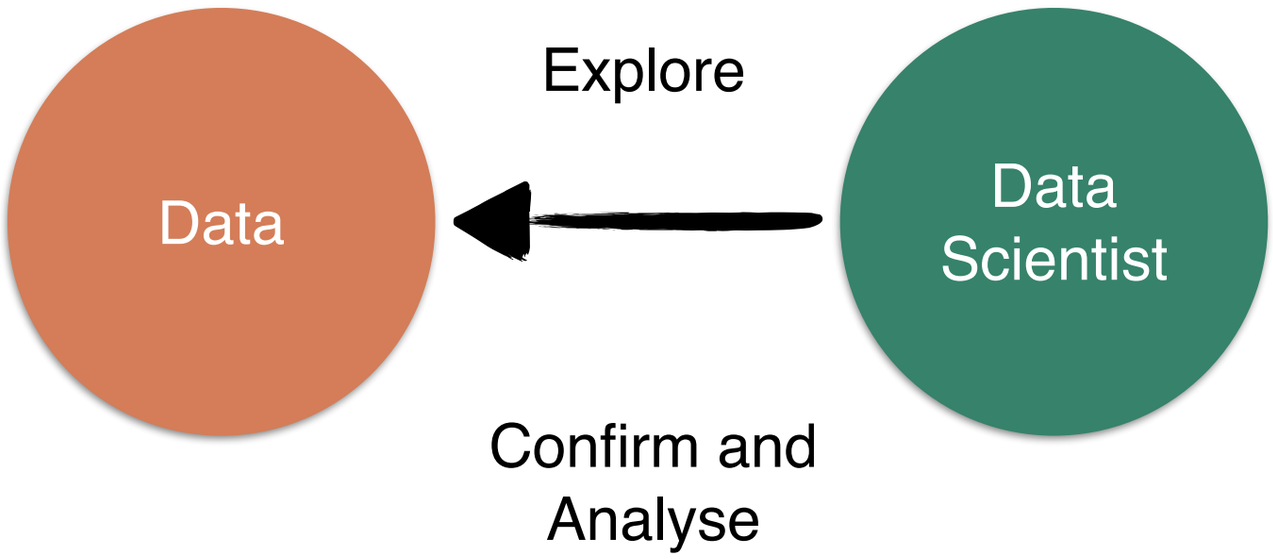


Communication & perception of results

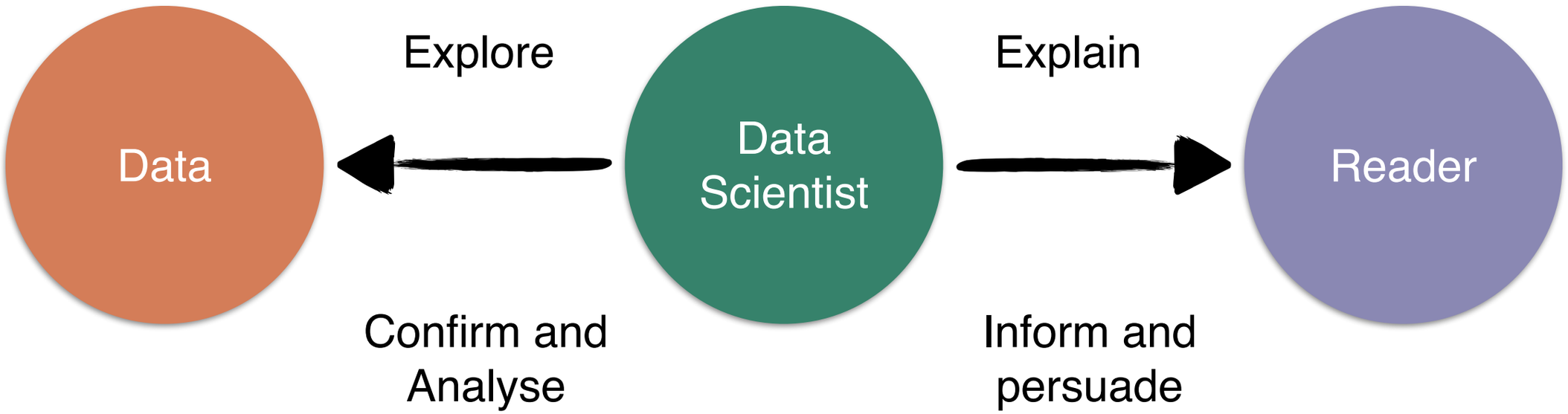
Combined in a meaningful way!

Exploratory vs Explanatory

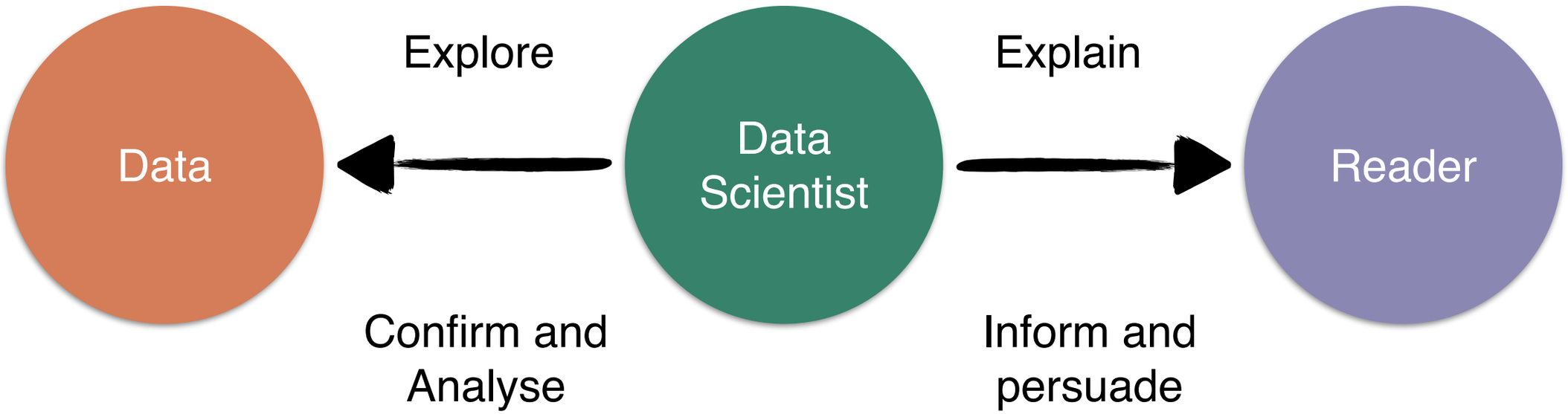
Exploratory vs Explanatory



Exploratory vs Explanatory



Exploratory vs Explanatory



Think about the audience!

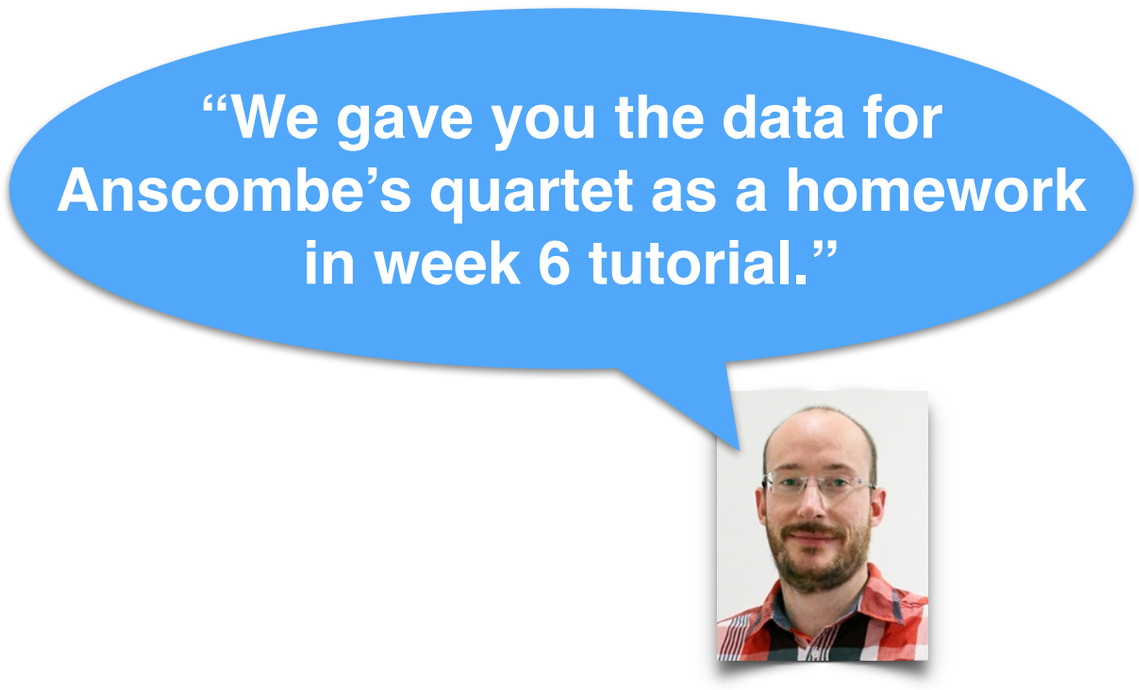
Why Data Vis Matters?

Why Data Vis Matters?

- Power of the visual medium

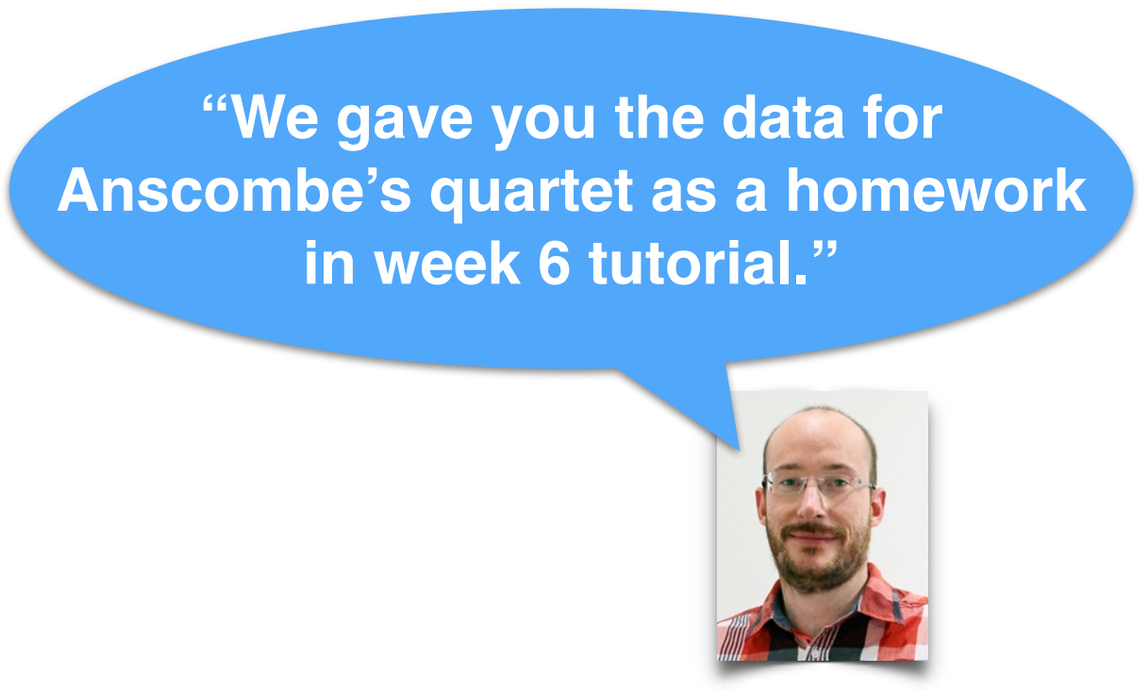
Why Data Vis Matters?

- Power of the visual medium
- ##Anscombe’s quartet



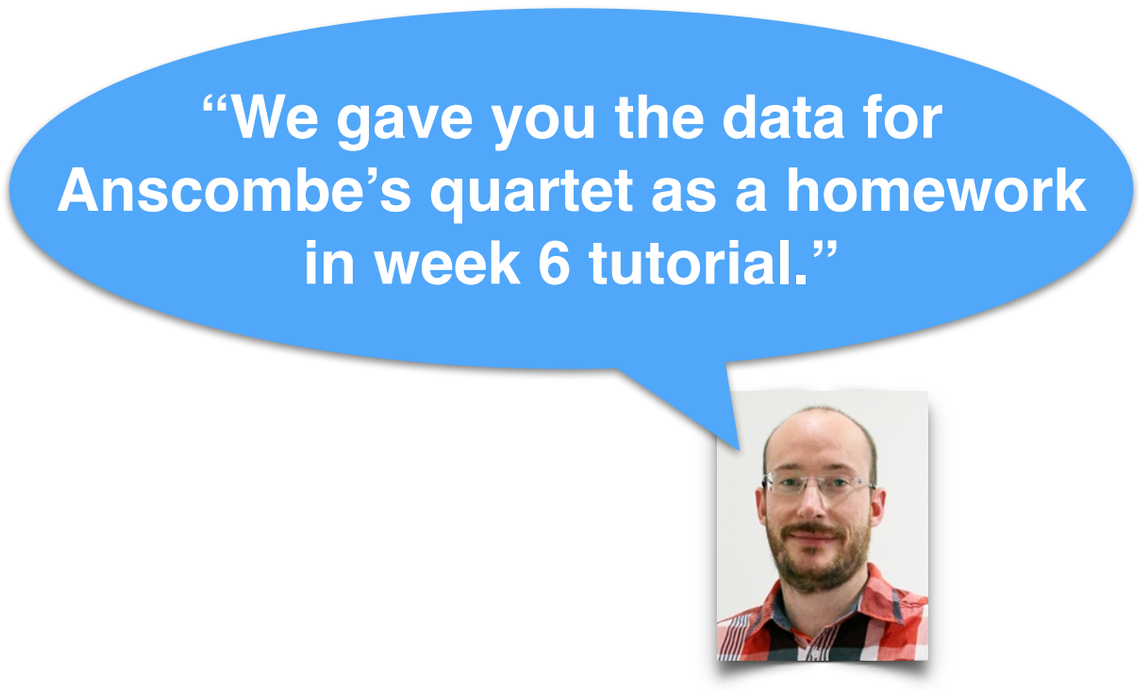
Why Data Vis Matters?

- Power of the visual medium
- ##Anscombe’s quartet
- Four datasets with identical mean, variance and correlation and linear regression line.

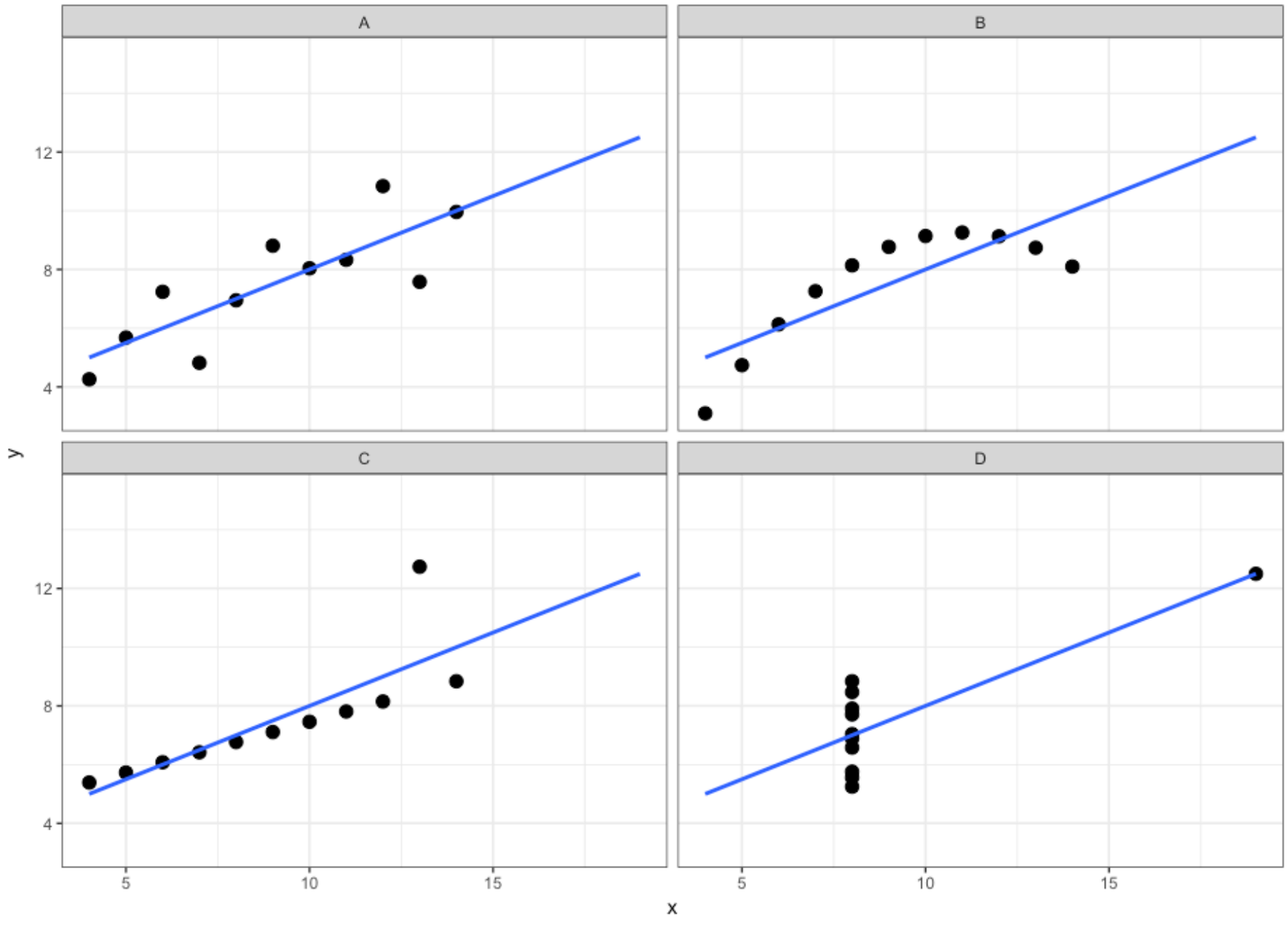


Why Data Vis Matters?

- Power of the visual medium
- ##Anscombe's quartet
- Four datasets with identical mean, variance and correlation and linear regression line.
- But…

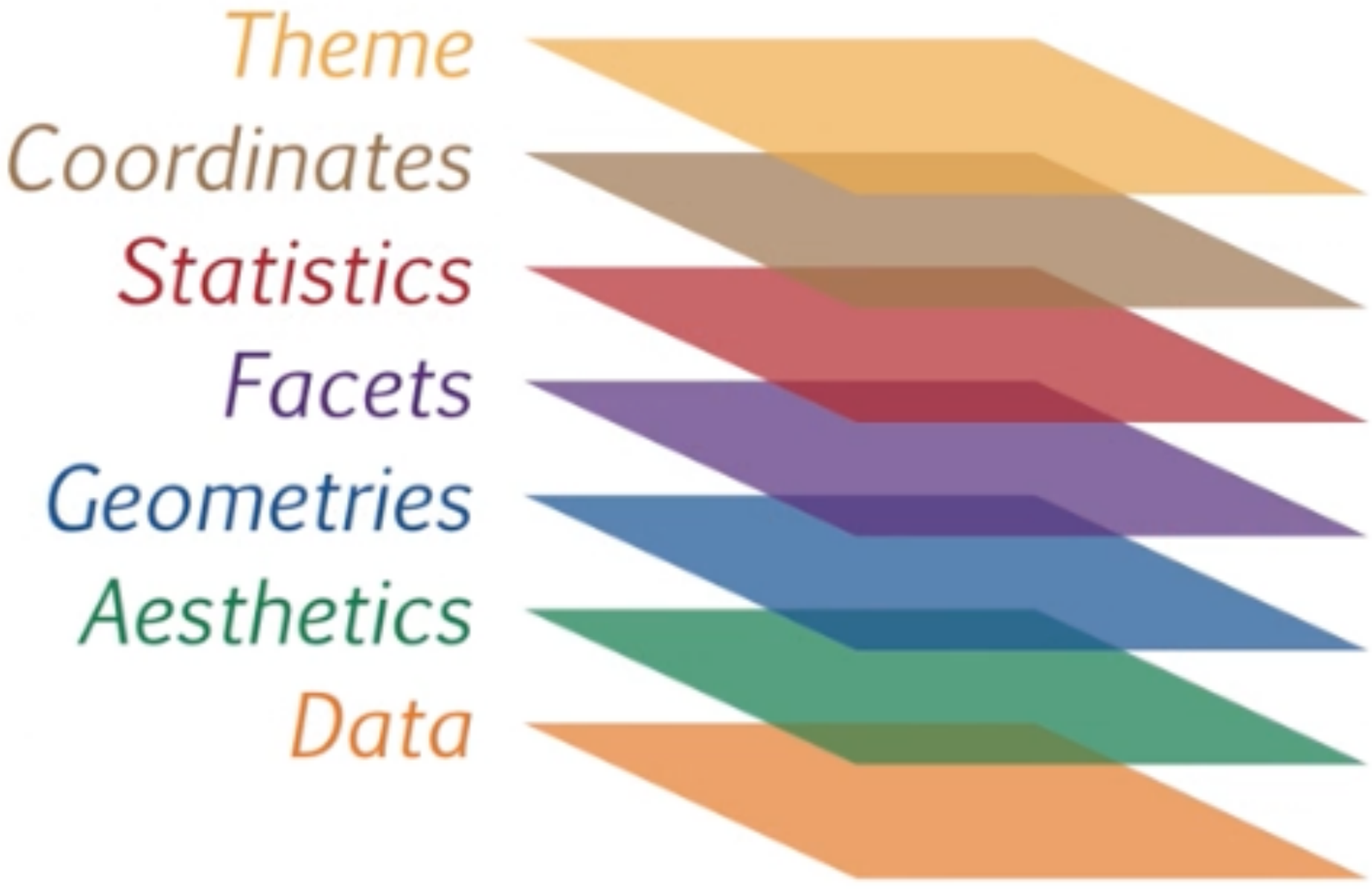


A difference ‘picture’

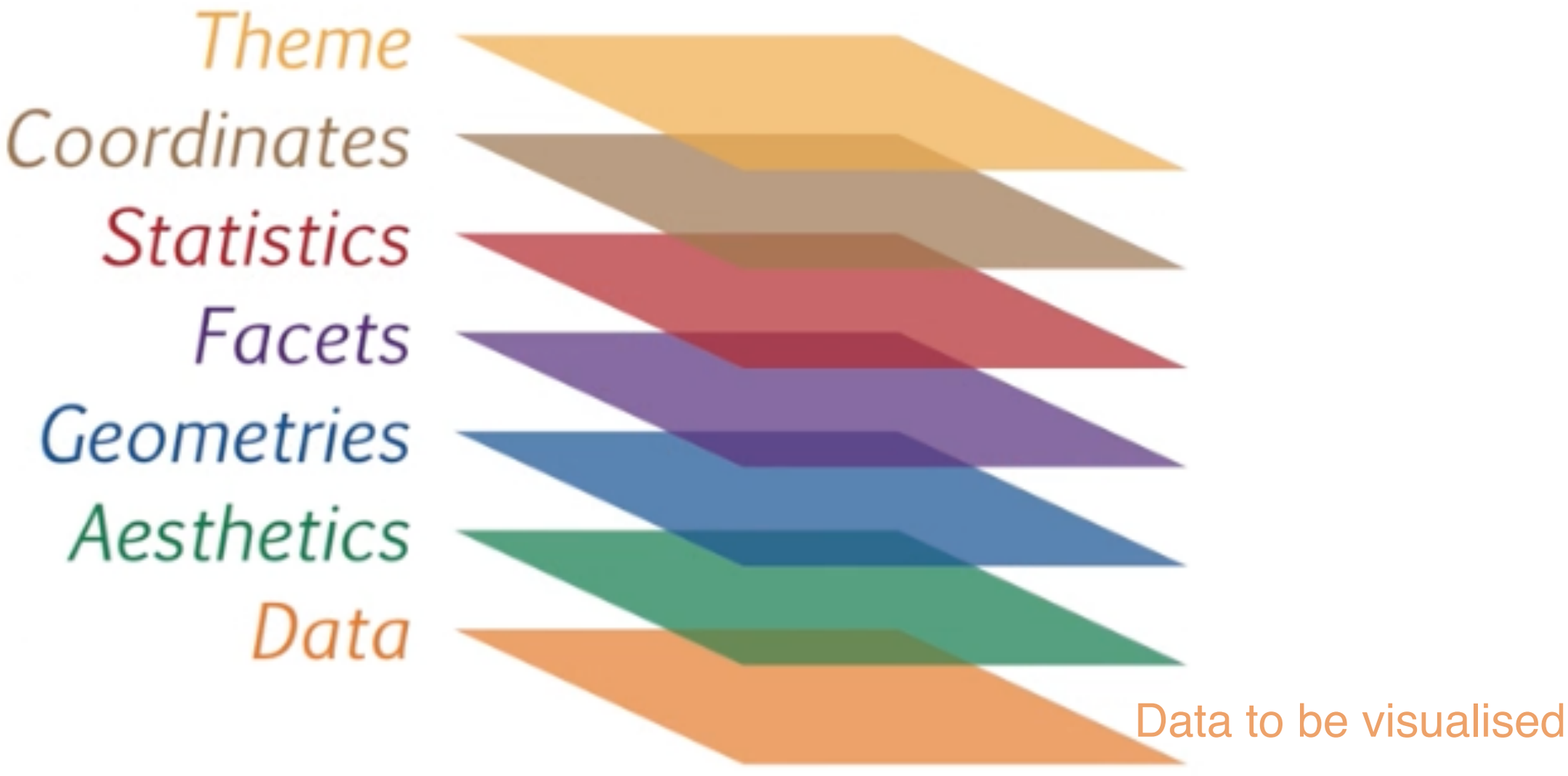


A Grammar of Graphics underpins data visualisation

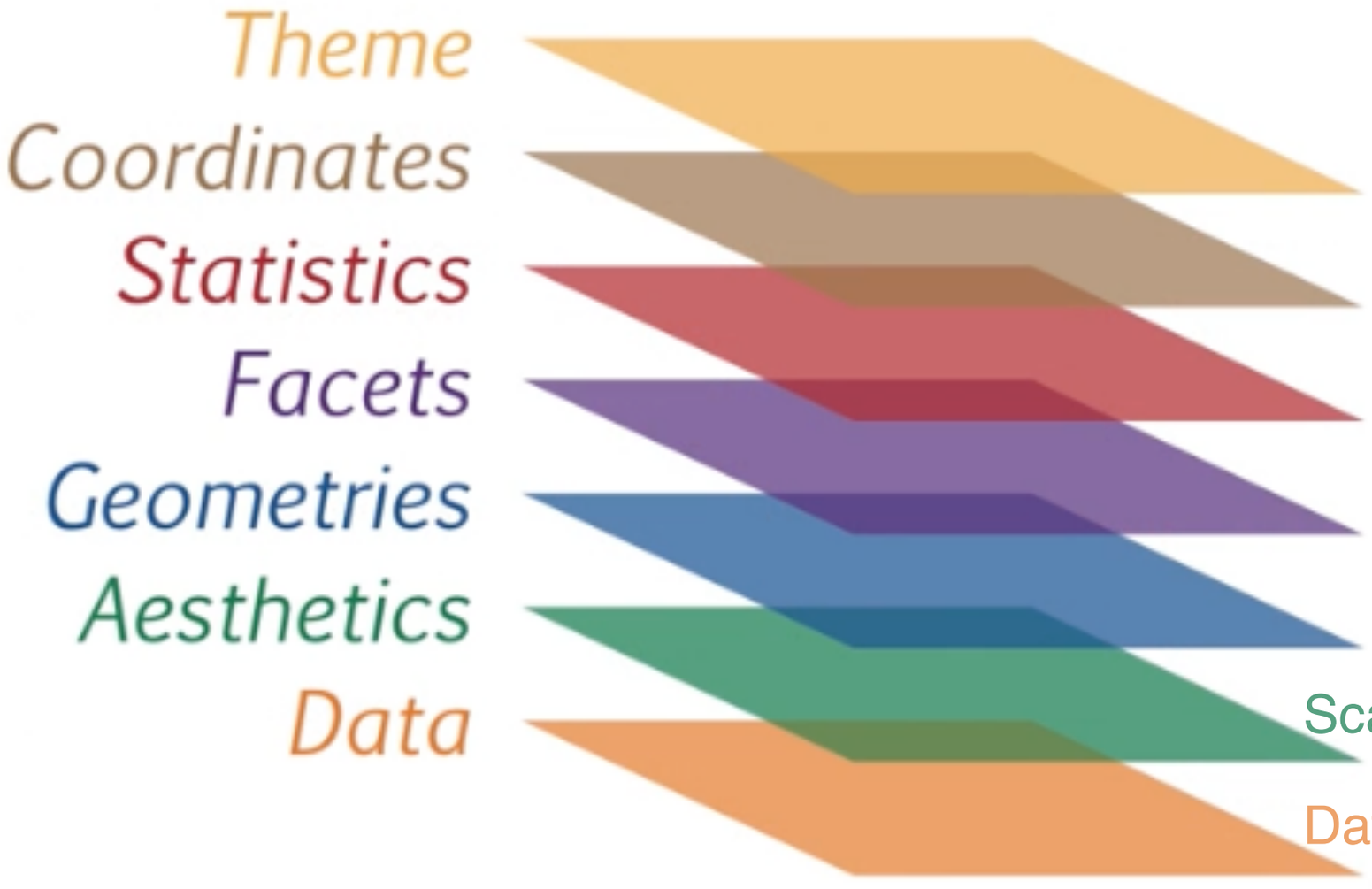
##Grammar of graphical elements



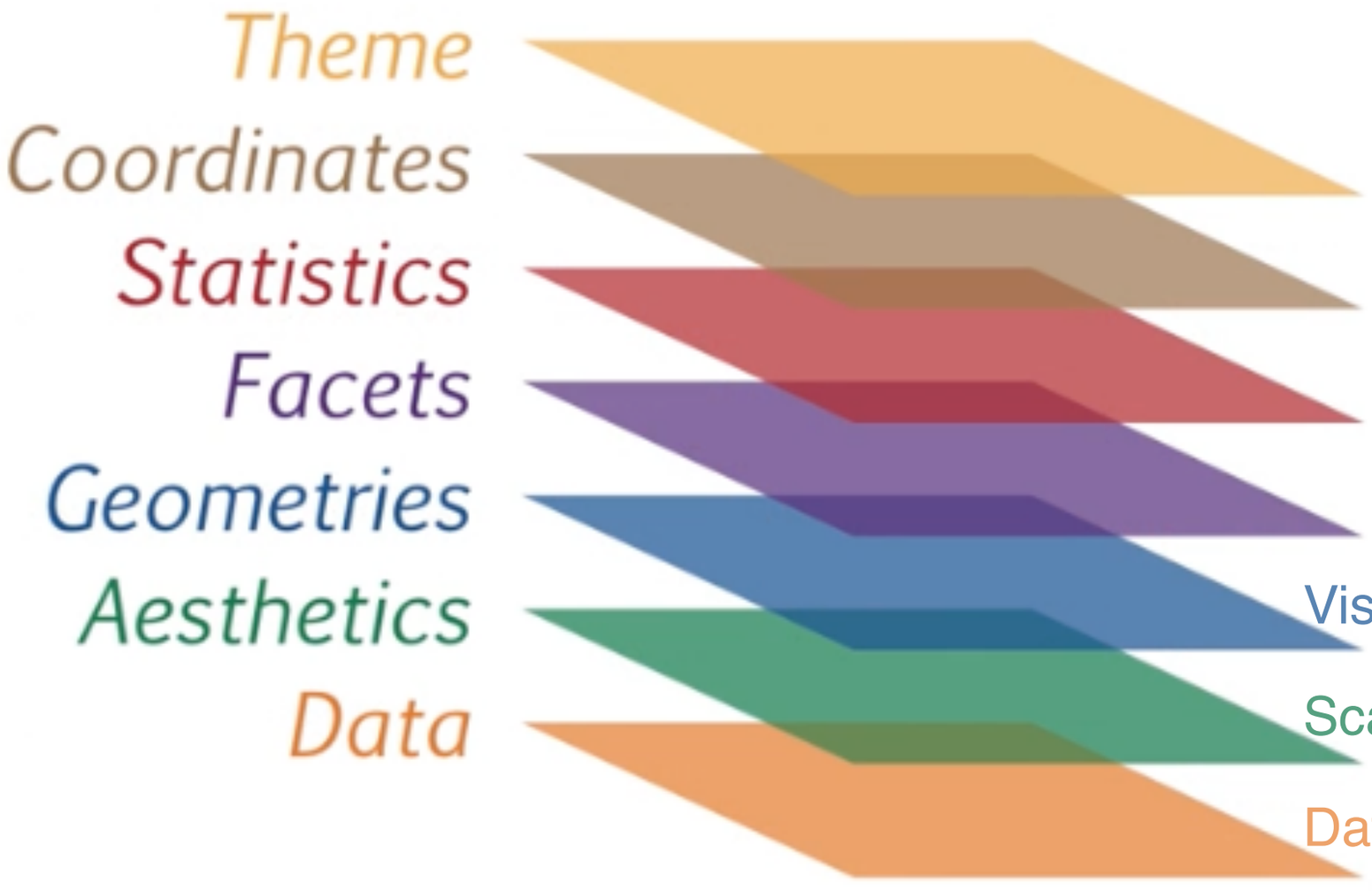
##Grammar of graphical elements



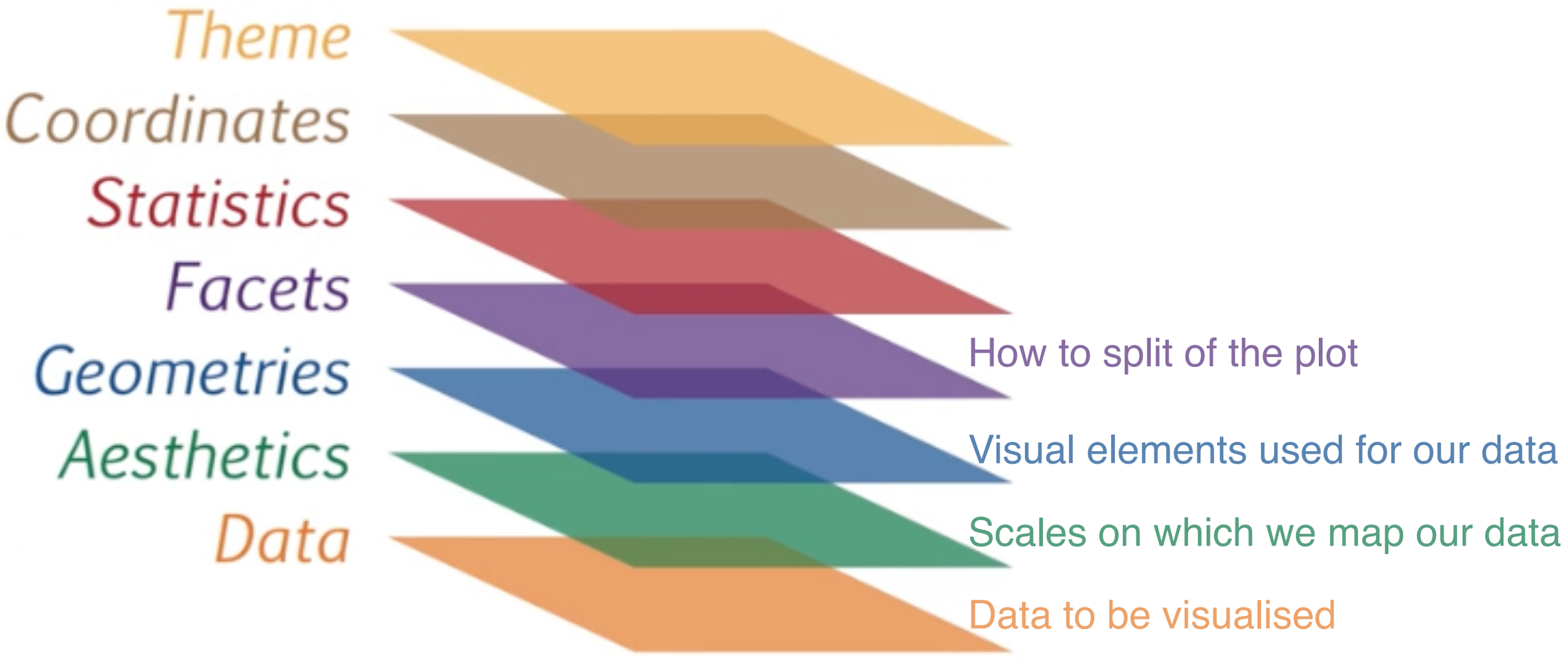
##Grammar of graphical elements



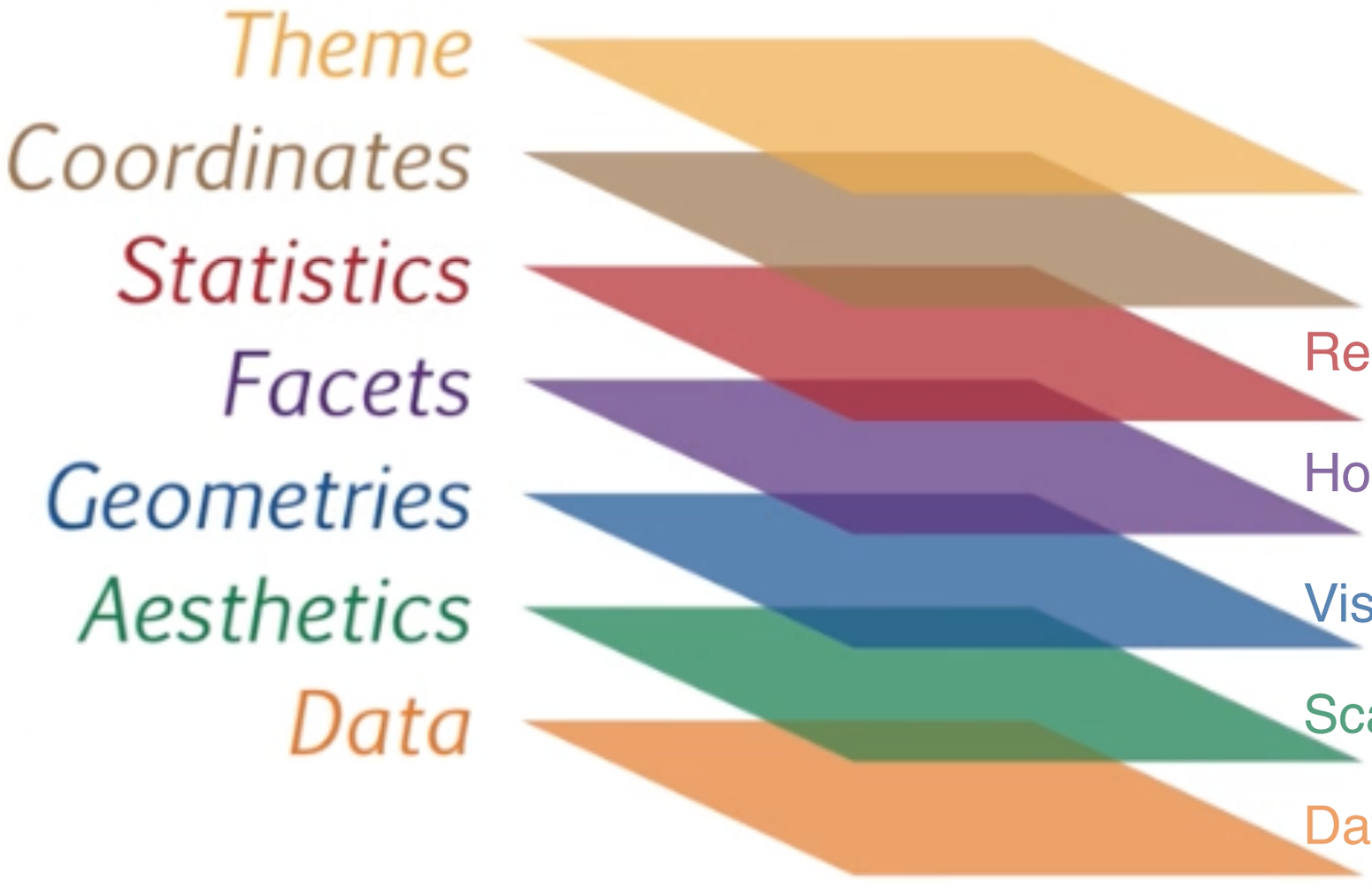
##Grammar of graphical elements



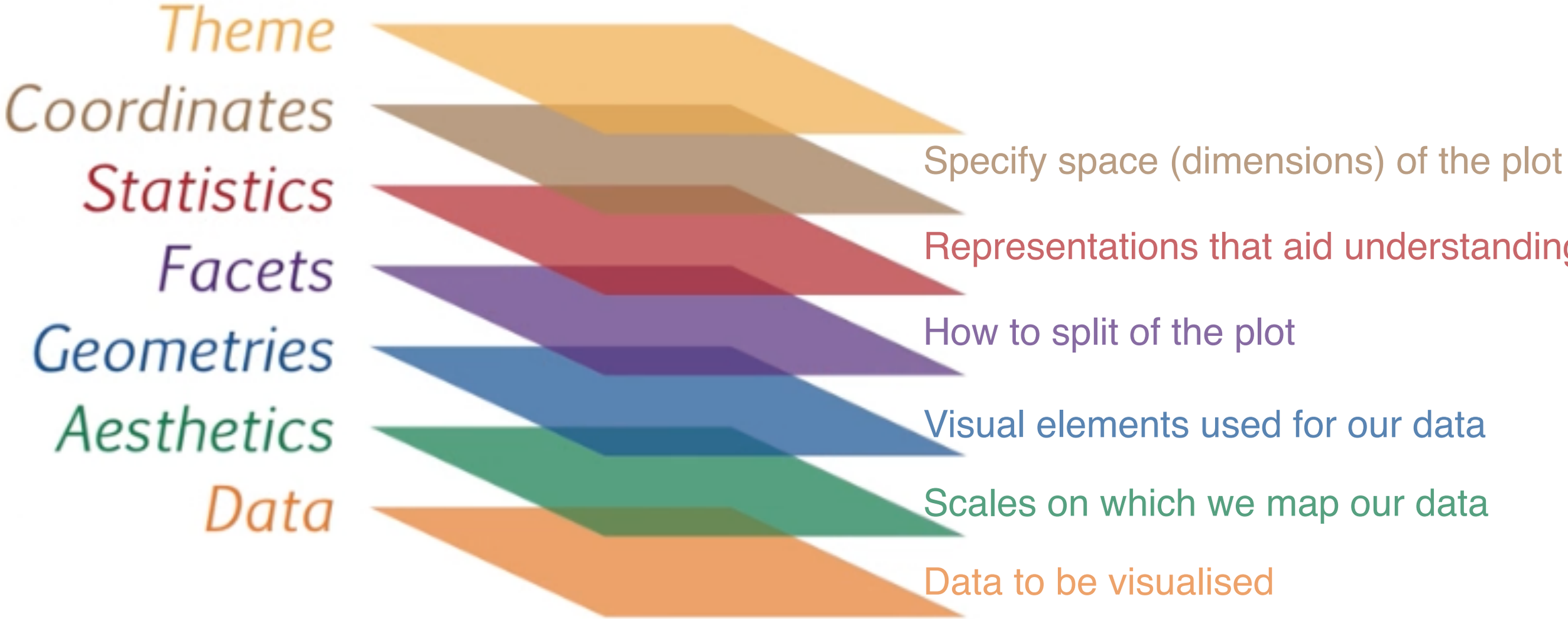
##Grammar of graphical elements



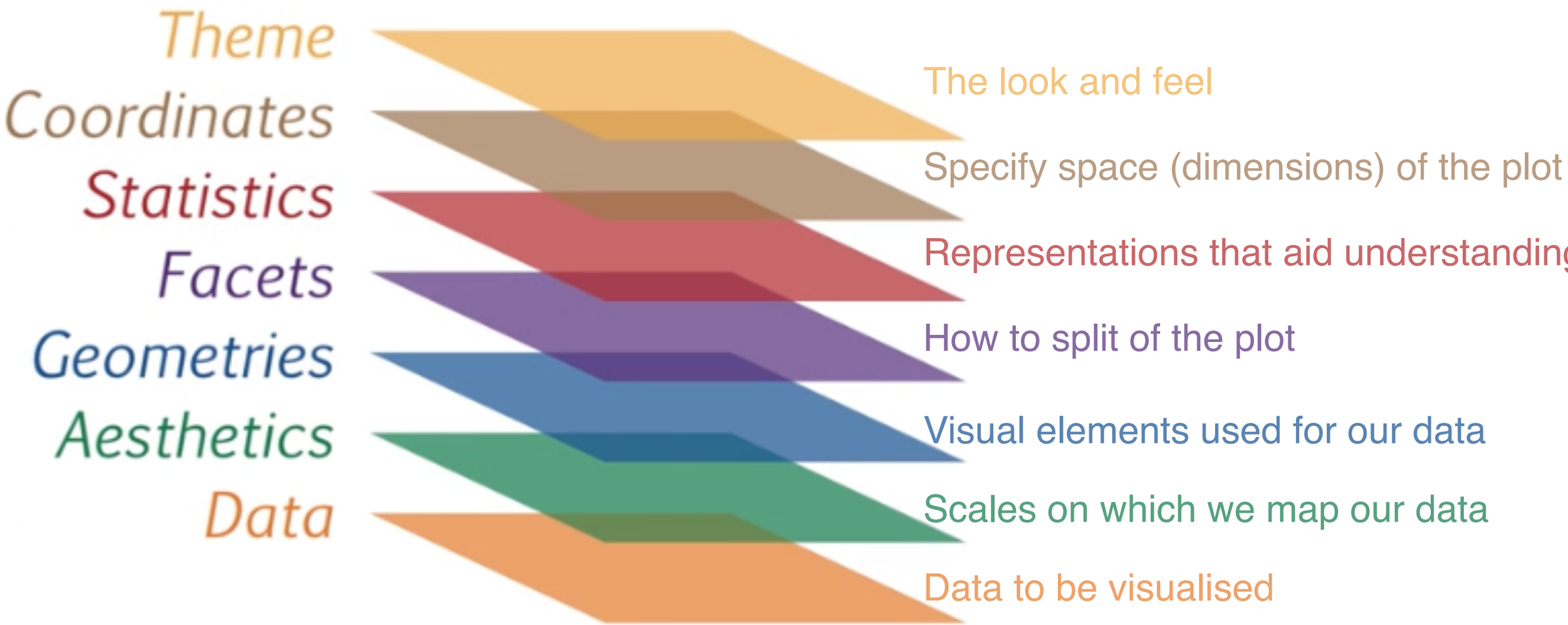
##Grammar of graphical elements



##Grammar of graphical elements



##Grammar of graphical elements



ggplot2 graphical elements

```
> ggplot(diamonds, aes(x=carat, y=price))
  + geom_point(alpha=.4)
  + facet_wrap(~clarity)
  + geom_smooth(method="lm", se=F)
  + scale_x_continuous(limits=c(0,3), breaks = seq(0,3,
  0.5)) + scale_y_continuous(labels = dollar)
   + xlab("Mass (carat)") + ylab("Price (US$)")
  + theme_minimal()
```

ggplot2 graphical elements

Data

```
> ggplot(diamonds, aes(x=carat, y=price))
  + geom_point(alpha=.4)
  + facet_wrap(~clarity)
  + geom_smooth(method="lm", se=F)
  + scale_x_continuous(limits=c(0,3), breaks = seq(0,3,
0.5)) + scale_y_continuous(labels = dollar)
   + xlab("Mass (carat)") + ylab("Price (US$)")
  + theme_minimal()
```

ggplot2 graphical elements

Data **Aesthetics**

```
> ggplot(diamonds, aes(x=carat, y=price))
  + geom_point(alpha=.4)
  + facet_wrap(~clarity)
  + geom_smooth(method="lm", se=F)
  + scale_x_continuous(limits=c(0,3), breaks = seq(0,3,
  0.5)) + scale_y_continuous(labels = dollar)
   + xlab("Mass (carat)") + ylab("Price (US$)")
  + theme_minimal()
```

ggplot2 graphical elements

Data Aesthetics

```
> ggplot(diamonds, aes(x=carat, y=price))
```

Geometry

```
+ geom_point(alpha=.4)
+ facet_wrap(~clarity)
+ geom_smooth(method="lm", se=F)
+ scale_x_continuous(limits=c(0,3), breaks = seq(0,3,
0.5)) + scale_y_continuous(labels = dollar)
 + xlab("Mass (carat)") + ylab("Price (US$)")
+ theme_minimal()
```

ggplot2 graphical elements

Data Aesthetics

```
> ggplot(diamonds, aes(x=carat, y=price))
```

Geometry

```
+ geom_point(alpha=.4)
```

Facets

```
 + facet_wrap(~clarity)
 + geom_smooth(method="lm", se=F)
 + scale_x_continuous(limits=c(0,3), breaks = seq(0,3,
0.5)) + scale_y_continuous(labels = dollar)
  + xlab("Mass (carat)") + ylab("Price (US$)")
 + theme_minimal()
```

ggplot2 graphical elements

Data Aesthetics

```
> ggplot(diamonds, aes(x=carat, y=price))
```

Geometry `+ geom_point(alpha=.4)`

Facets `+ facet_wrap(~clarity)`

Statistics `+ geom_smooth(method="lm", se=F)`

```
+ scale_x_continuous(limits=c(0,3), breaks = seq(0,3,
0.5)) + scale_y_continuous(labels = dollar)
 + xlab("Mass (carat)") + ylab("Price (US$)")
+ theme_minimal()
```

ggplot2 graphical elements

Data Aesthetics

```
> ggplot(diamonds, aes(x=carat, y=price))
```

Geometry `+ geom_point(alpha=.4)`

Facets `+ facet_wrap(~clarity)`

Statistics `+ geom_smooth(method="lm", se=F)`

Coordinates `+ scale_x_continuous(limits=c(0,3), breaks = seq(0,3,`

`0.5)) + scale_y_continuous(labels = dollar)`

`+ xlab("Mass (carat)") + ylab("Price (US$)")`

`+ theme_minimal()`

ggplot2 graphical elements

Data Aesthetics

```
> ggplot(diamonds, aes(x=carat, y=price))
```

Geometry `+ geom_point(alpha=.4)`

Facets `+ facet_wrap(~clarity)`

Statistics `+ geom_smooth(method="lm", se=F)`

Coordinates `+ scale_x_continuous(limits=c(0,3), breaks = seq(0,3,`
`0.5)) + scale_y_continuous(labels = dollar)`
`+ xlab("Mass (carat)") + ylab("Price (US$)")`

Theme `+ theme_minimal()`

ggplot2 graphical elements

```
> ggplot(
  + geom_p
  + facet_
  + geom_s
  + scale_
    0.5)) +
    + xlab(
  + theme_
```

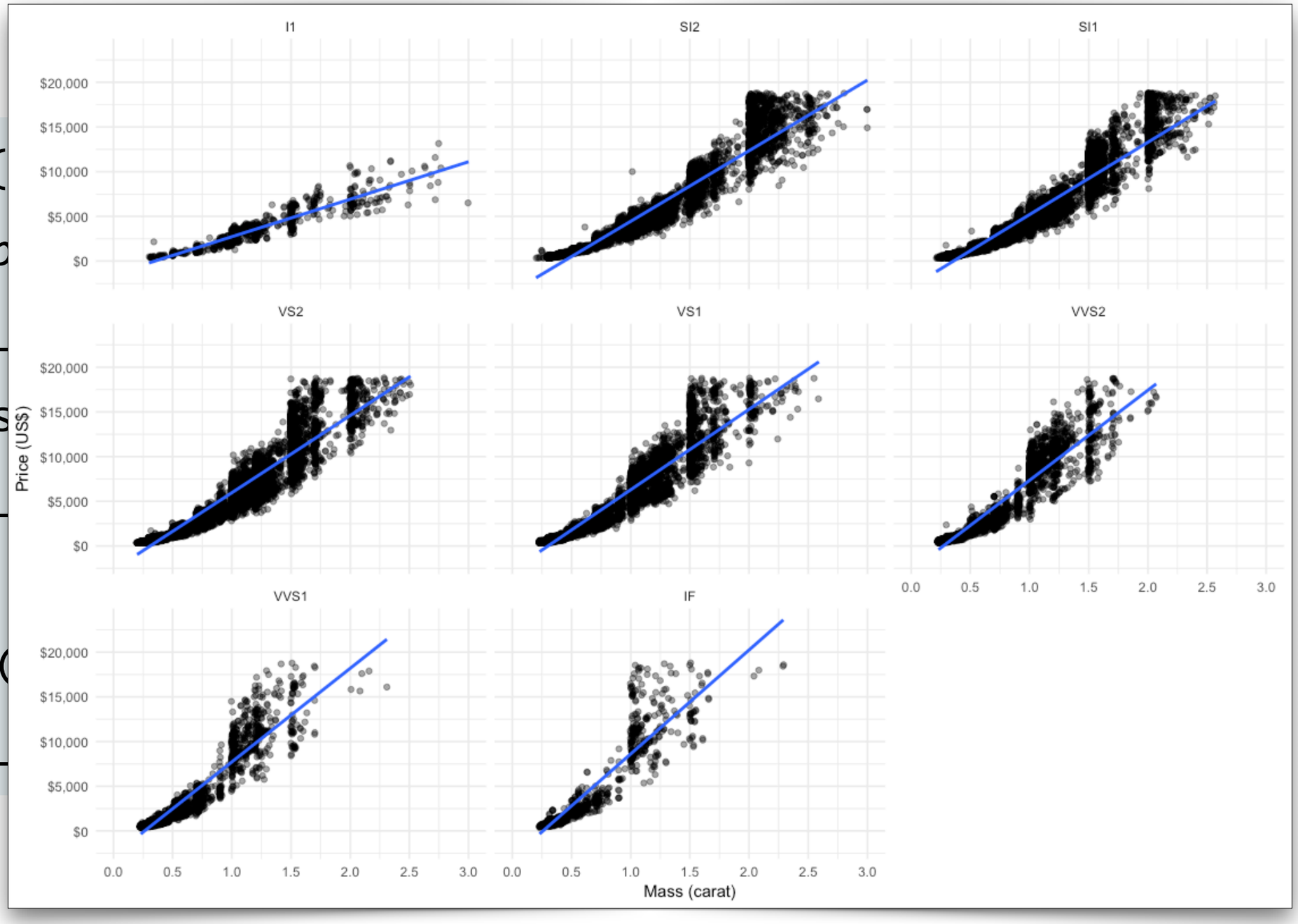
Geometry

Facets

Statistics

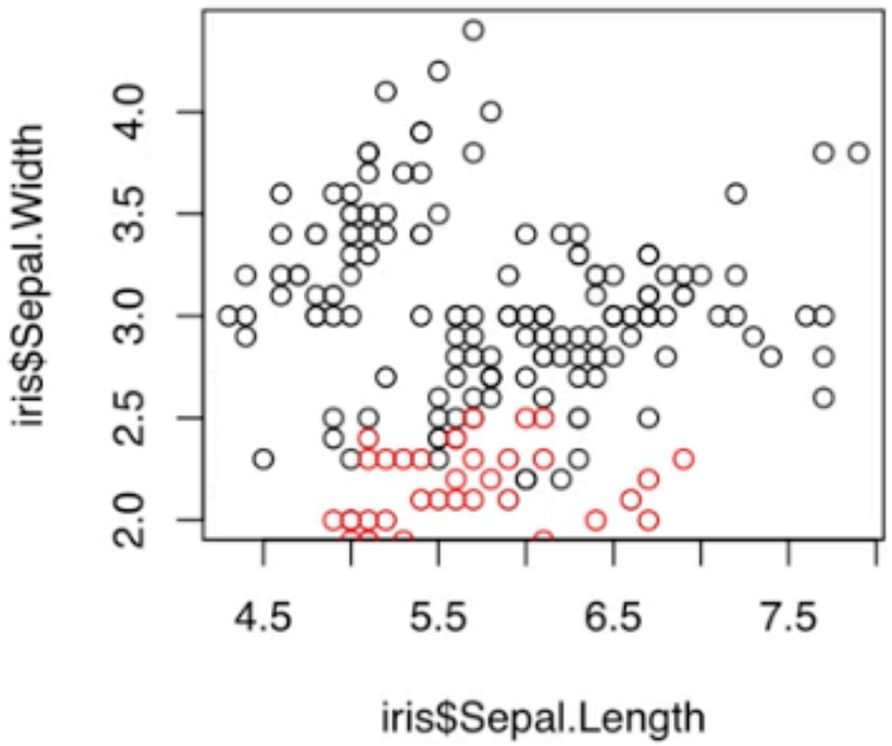
Coordinates

Theme



Base plotting package

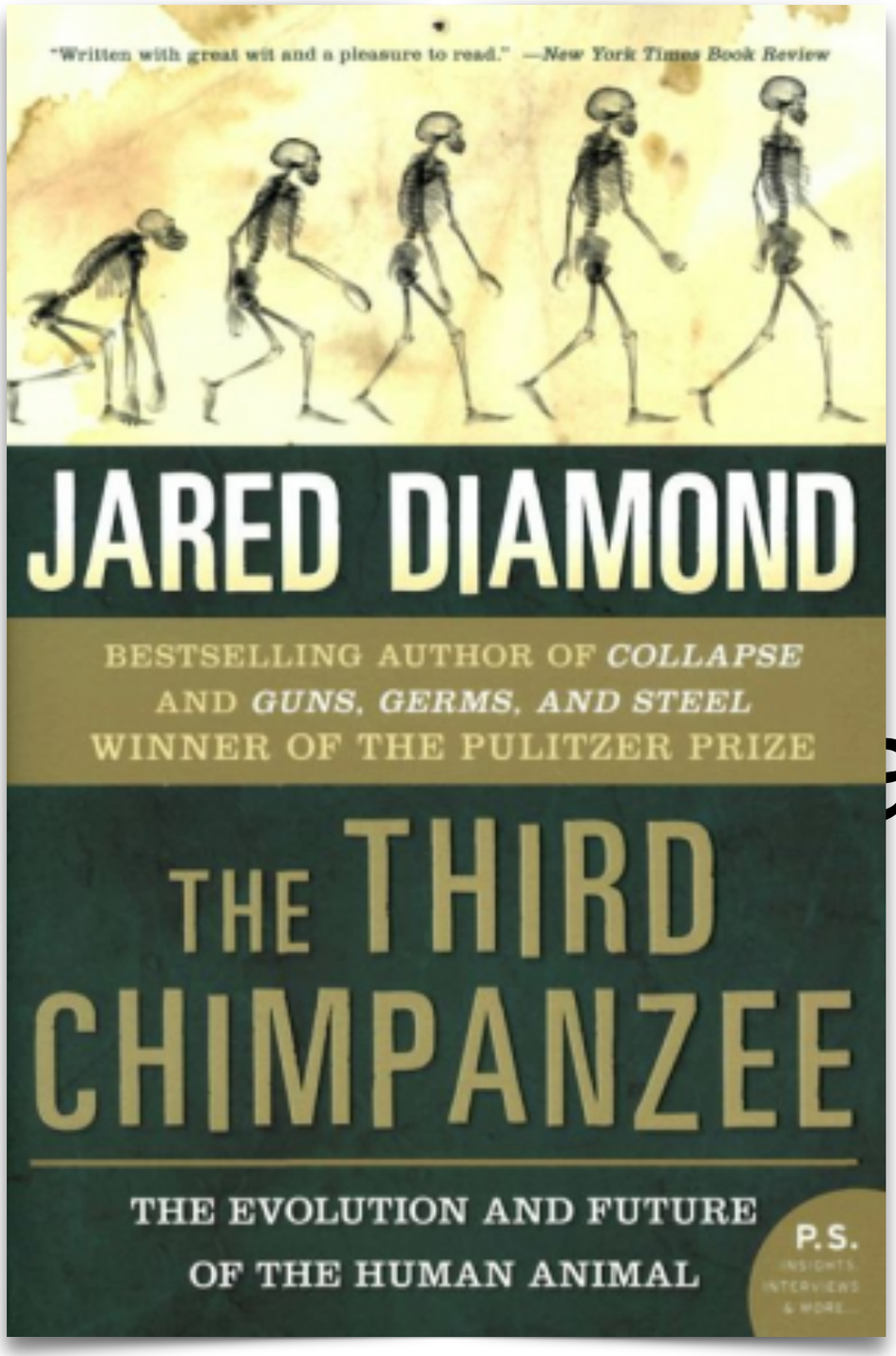
```
> plot(iris$Sepal.Length, iris$Sepal.Width)
> points(iris$Petal.Length, iris$Petal.Width, col = "red")
```



Limitations

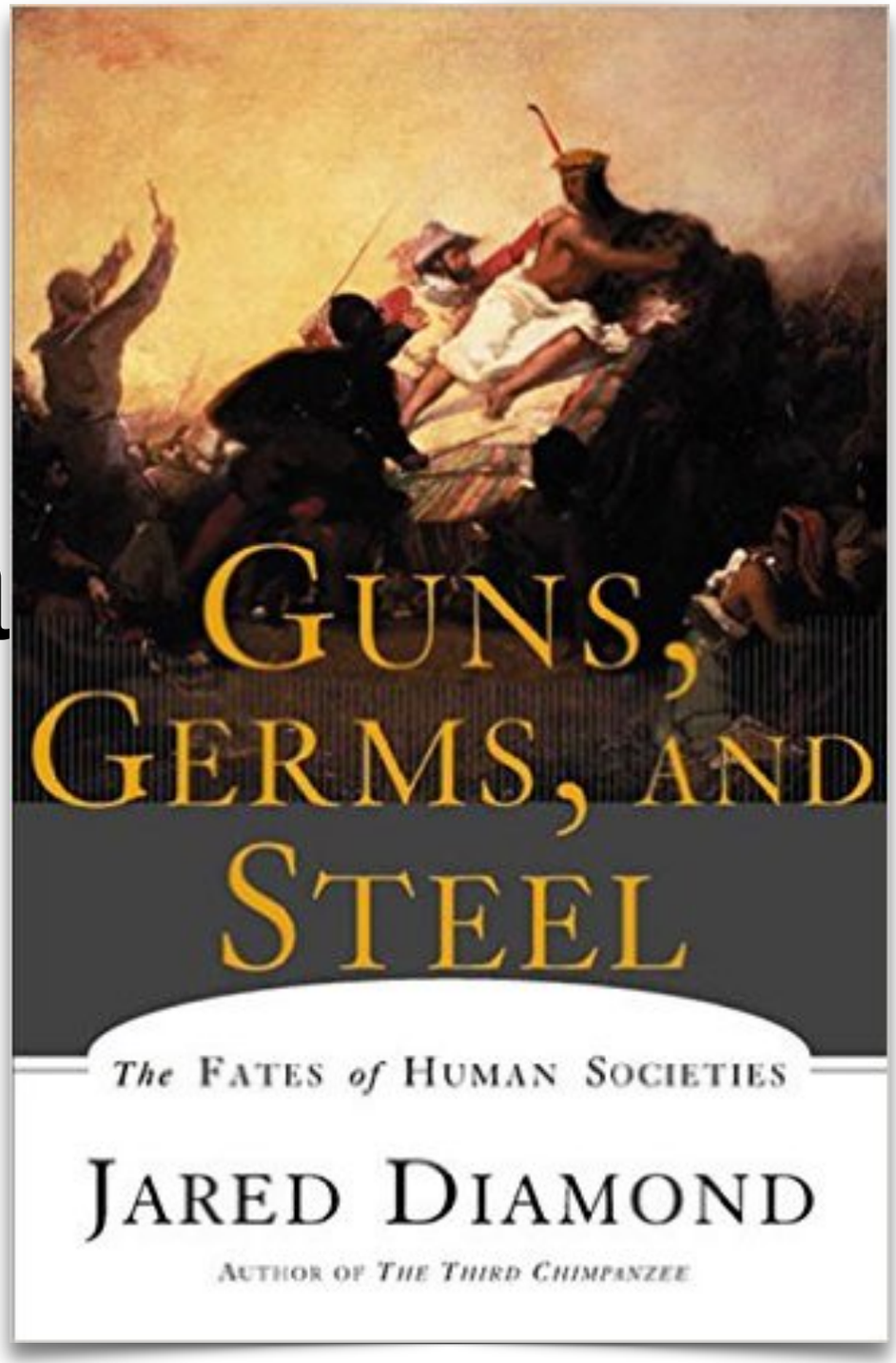
1. Plot doesn't get redrawn
2. Plot is drawn as an image
3. Need to manually add legend
4. No unified framework for plotting

Choosing appropriate visual cues


"Written with great wit and a pleasure to read." —New York Times Book Review
JARED DIAMOND
BESTSELLING AUTHOR OF COLLAPSE
AND GUNS, GERMS, AND STEEL
WINNER OF THE PULITZER PRIZE
THE THIRD CHIMPANZEE
THE EVOLUTION AND FUTURE
OF THE HUMAN ANIMAL
P.S.

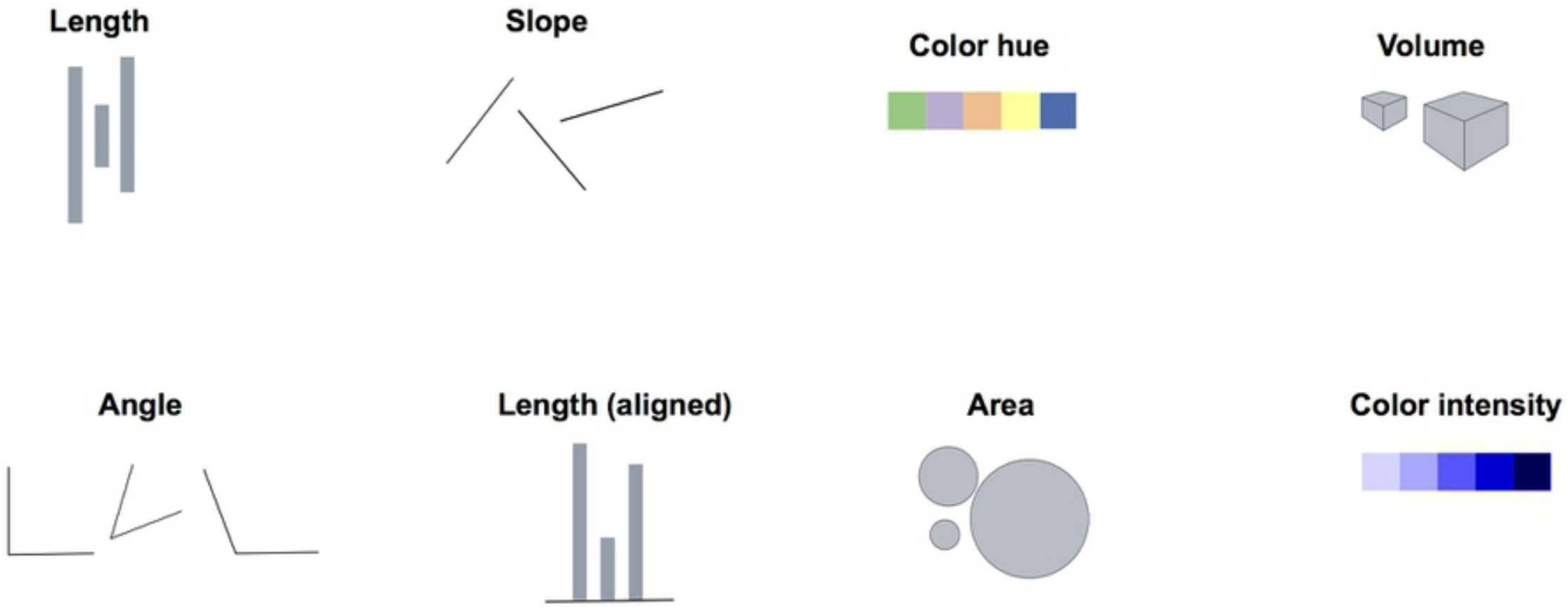
a

al

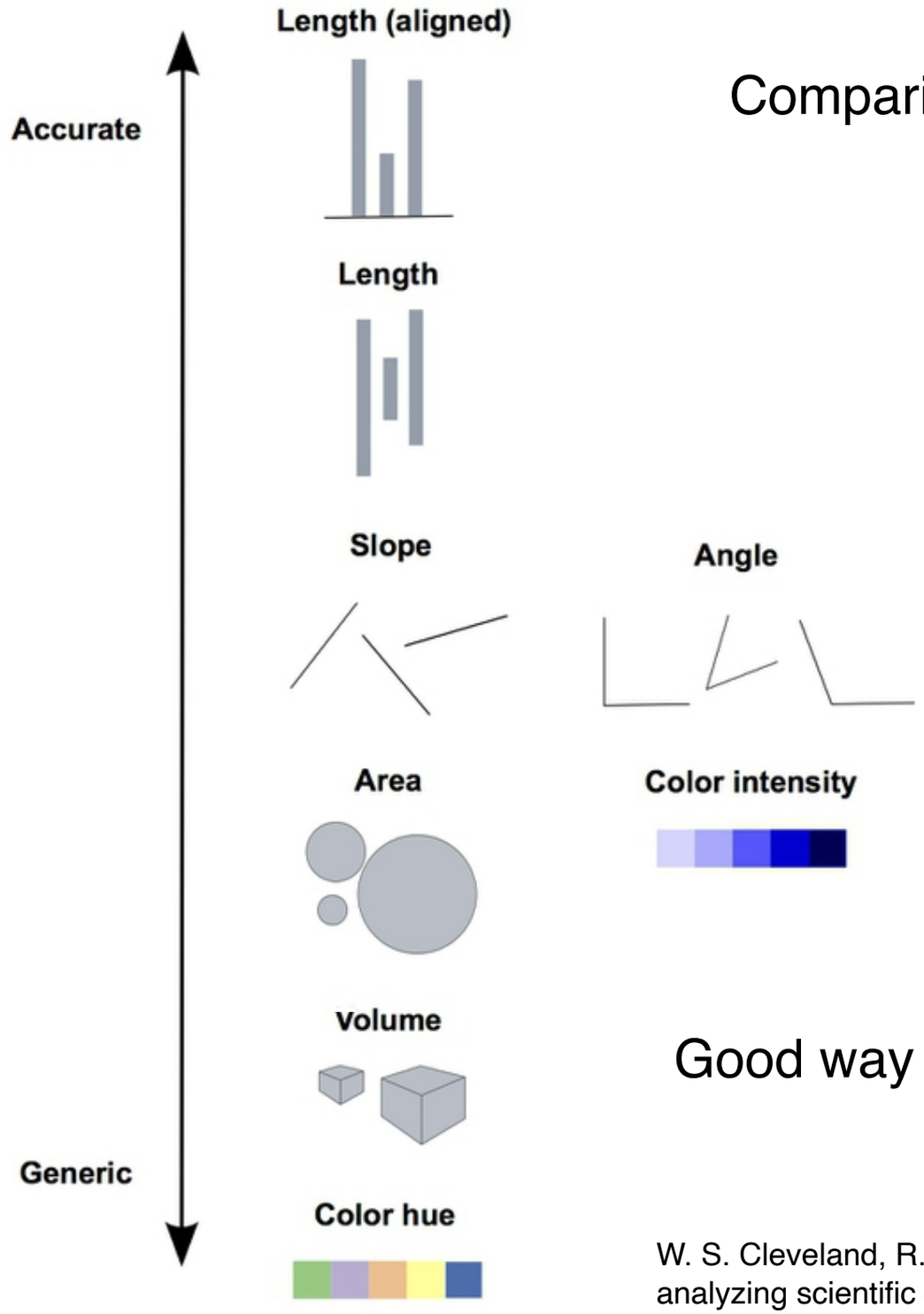

GUNS,
GERMS, AND
STEEL
The Fates of Human Societies
JARED DIAMOND
Author of The Third Chimpanzee

Visual Cues

Can you order these according to accuracy?



Accuracy for perceiving quantitative information

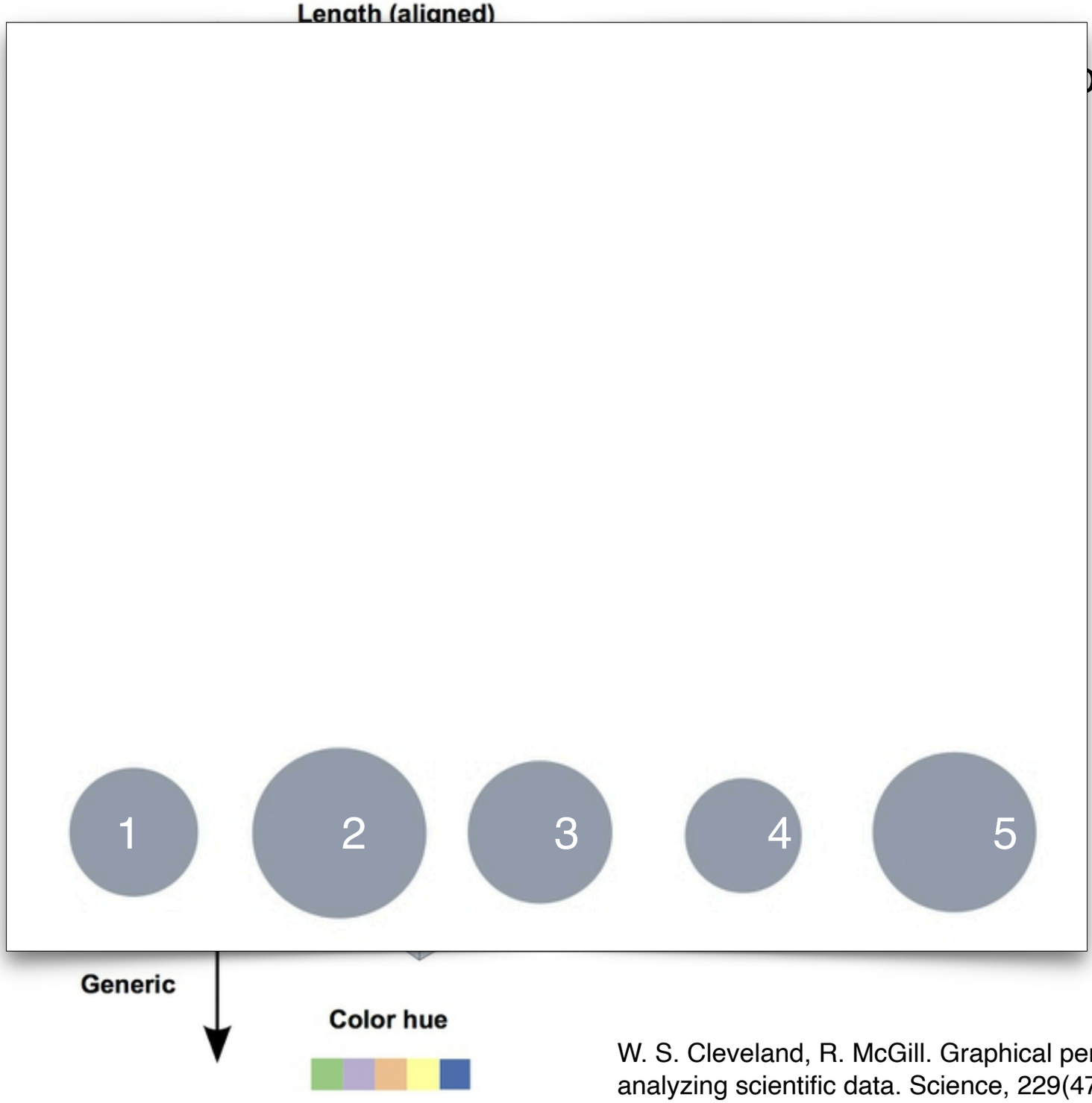


Comparisons with continuous variables

Good way of encoding categorical data

W. S. Cleveland, R. McGill. Graphical perception and graphical methods for analyzing scientific data. Science, 229(4716):828–833, 1985.

Accuracy for perceiving quantitative information

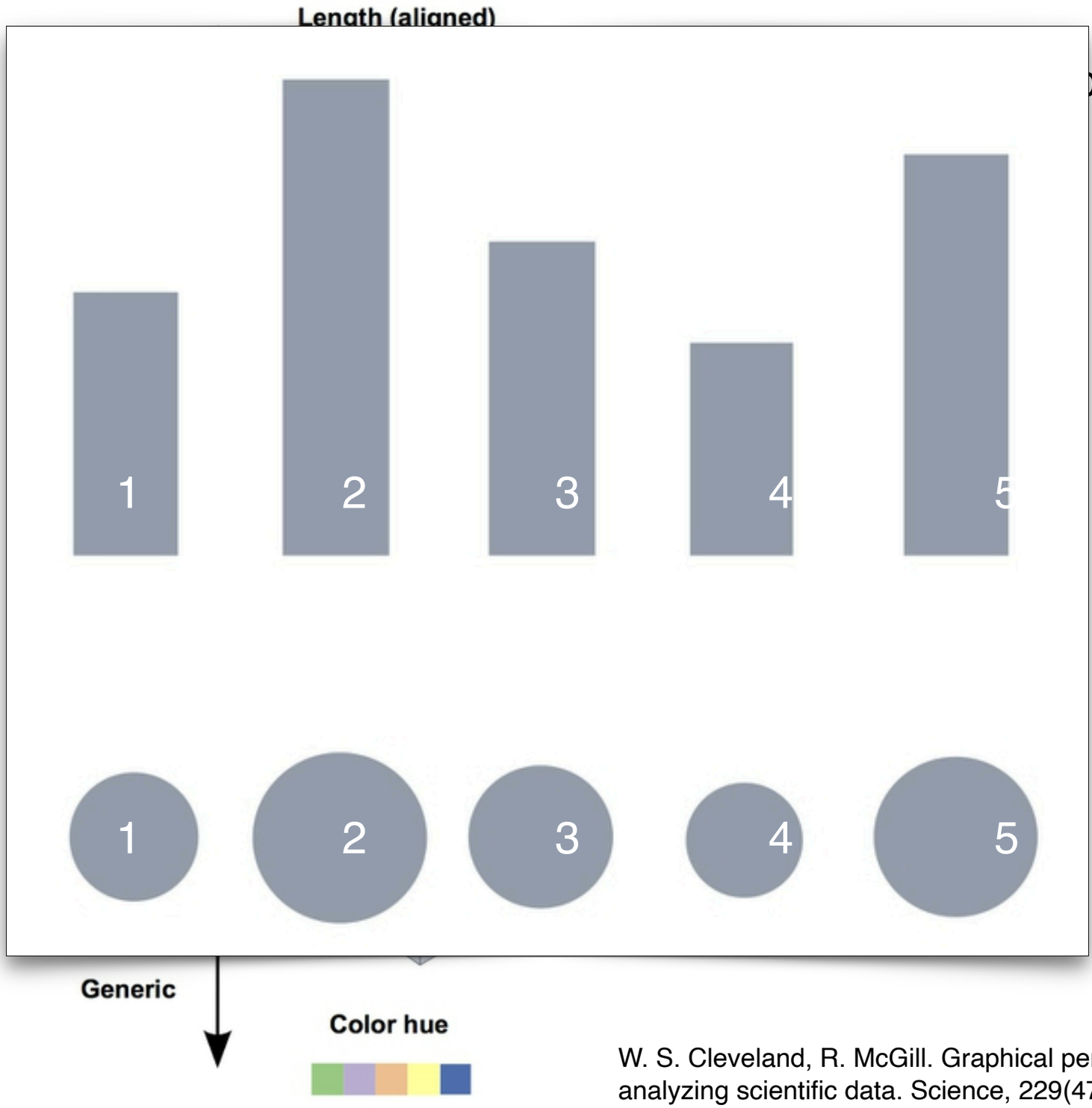


ntinuous variables

categorical data

W. S. Cleveland, R. McGill. Graphical perception and graphical methods for analyzing scientific data. Science, 229(4716):828–833, 1985.

Accuracy for perceiving quantitative information

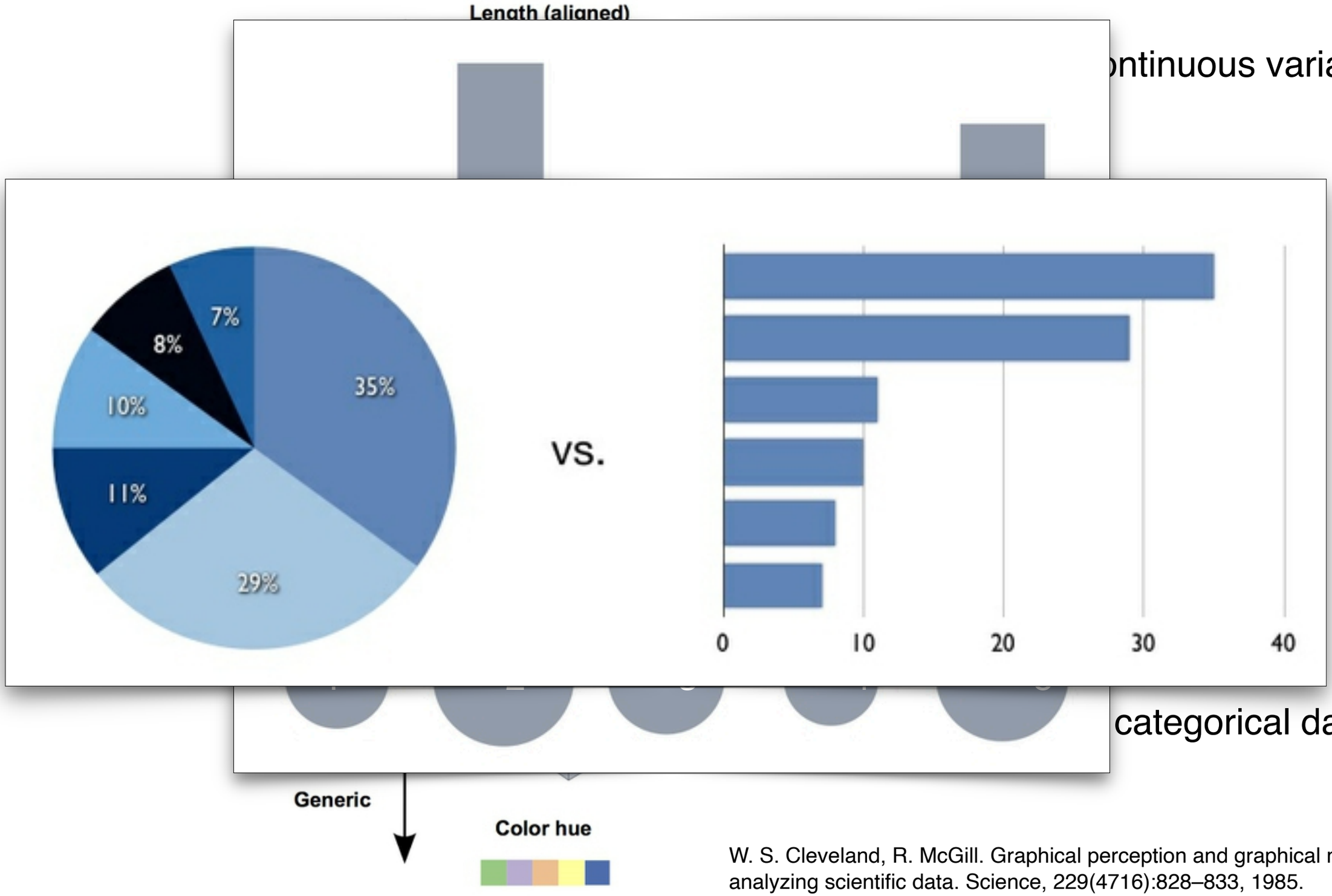


ntinuous variables

categorical data

W. S. Cleveland, R. McGill. Graphical perception and graphical methods for analyzing scientific data. Science, 229(4716):828–833, 1985.

Accuracy for perceiving quantitative information



ntinuous variables

categorical data

W. S. Cleveland, R. McGill. Graphical perception and graphical methods for analyzing scientific data. Science, 229(4716):828–833, 1985.

Chart Suggestions—A Thought-Starter

What would you like to show?

Comparison

Relationship

- Two Variables
  - Scatter Chart
  - Bubble Chart
- Three Variables
  - Bubble Chart

Distribution

- Single Variable
  - Few Data Points
    - Column Histogram
  - Many Data Points
    - Line Histogram
- Two Variables
  - Scatter Chart
- Three Variables
  - 3D Area Chart

Composition

- Changing Over Time
  - Few Periods
    - Only Relative Differences Matter
      - Stacked 100% Column Chart
    - Relative and Absolute Differences Matter
      - Stacked Column Chart
  - Many Periods
    - Only Relative Differences Matter
      - Stacked 100% Area Chart
    - Relative and Absolute Differences Matter
      - Stacked Area Chart
- Static
  - Simple Share of Total
    - Pie Chart
  - Accumulation or Subtraction to Total
    - Waterfall Chart
  - Components of Components
    - Stacked 100% Column Chart with Subcomponents

Chart Selection Criteria

- Two Variables per Item
  - Among Items
    - One Variable per Item
      - Many Categories
        - Table or Table with Embedded Charts
      - Few Categories
        - Many Items
          - Bar Chart
        - Few Items
          - Column Chart
- Over Time
  - Many Periods
    - Cyclical Data
      - Circular Area Chart
    - Non-Cyclical Data
      - Line Chart
  - Few Periods
    - Single or Few Categories
      - Column Chart
    - Many Categories
      - Line Chart

www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com

Good Data Visualisation Practice

Reproducible research - RStudio & RMarkdown

- Document your steps

- No manual ‘one-offs’ (grep, sed)

- Share your working

RPubs http://rpubs.com/bevankoopman/266822

“We have used RMarkdown in last week computer lab.”

Reproducible research - RStudio & RMarkdown

- Document your steps

- No manual ‘one-offs’ (grep, sed)

- Share your working

RPubs http://rpubs.com/bevankoopman/266822

“We have used RMarkdown in last week computer lab.”

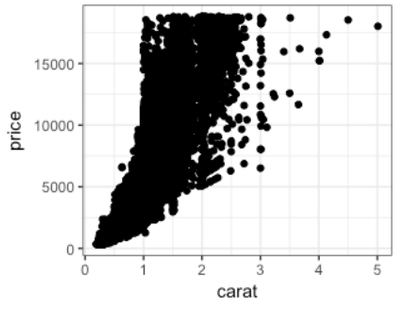
“Remember: you can integrate system calls (to grep, sed, etc) in RMarkdown.”

Data that supports Visualisation

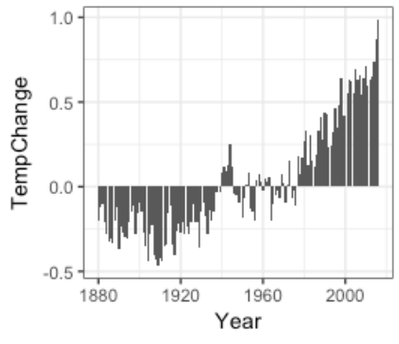
- Structure your data to support visualisation
  - Use meaningful, ‘displayable’ column names
  - Use ‘tidy’ data format:
    - Rows are observation, columns are variables (‘long’ rather than ‘wide’ formats)
    - Transform with `melt` function from `reshape2` library

##Different Plot Types

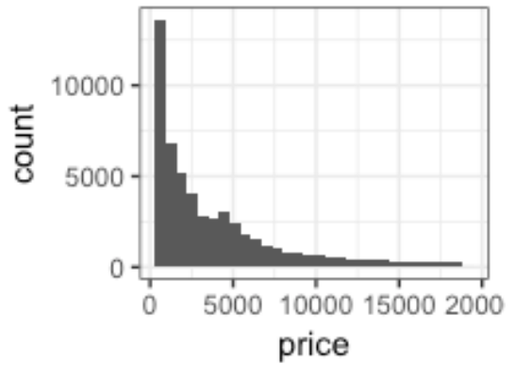
Scatter



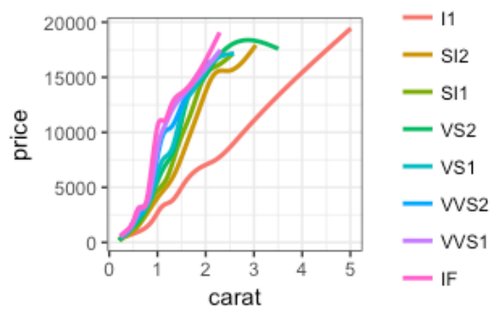
Bar plots



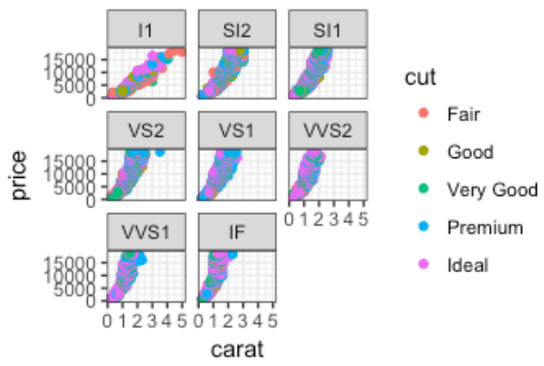
Histograms



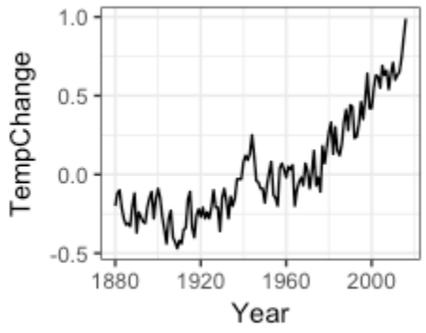
Trends



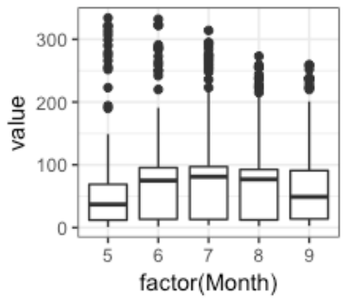
Facets



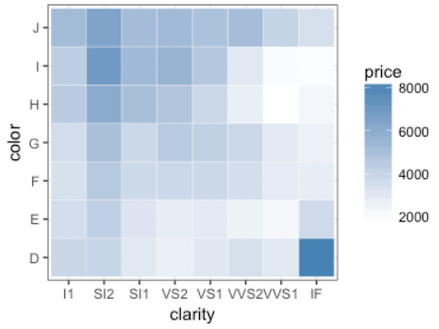
Line plots



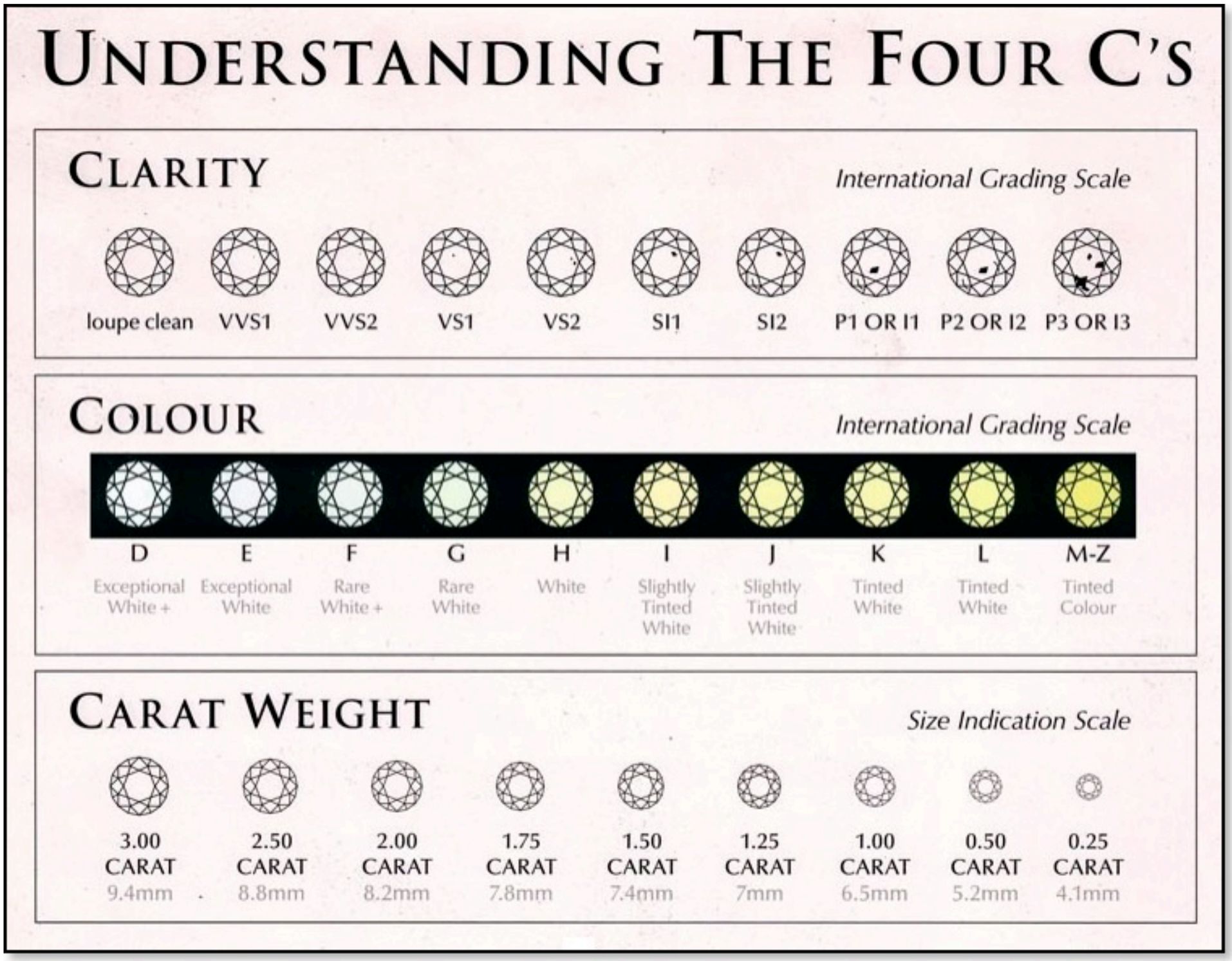
Box plots



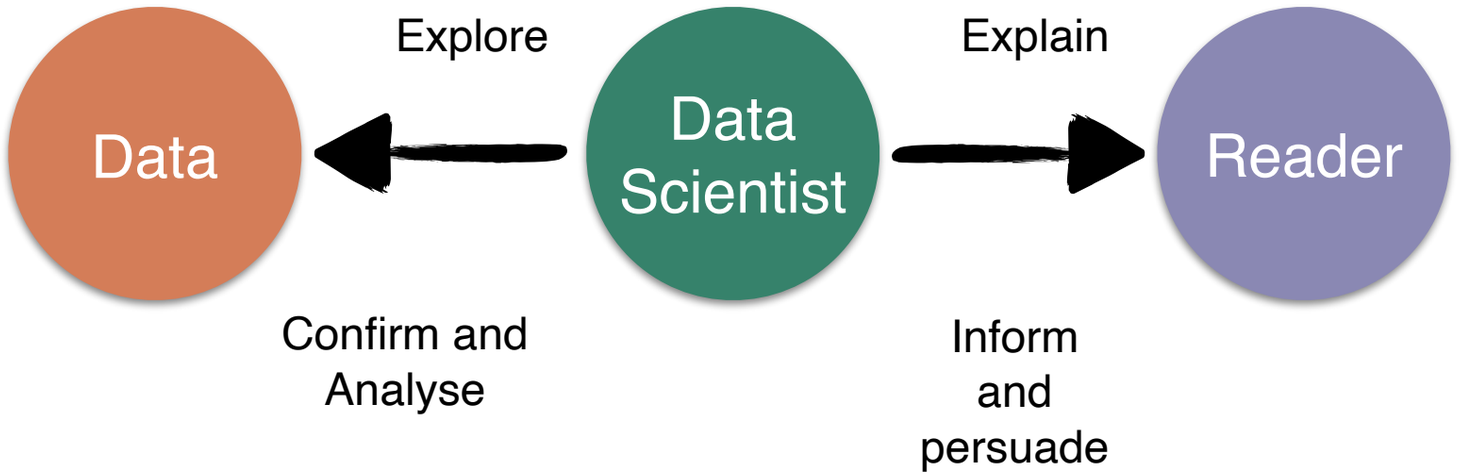
Heatmaps

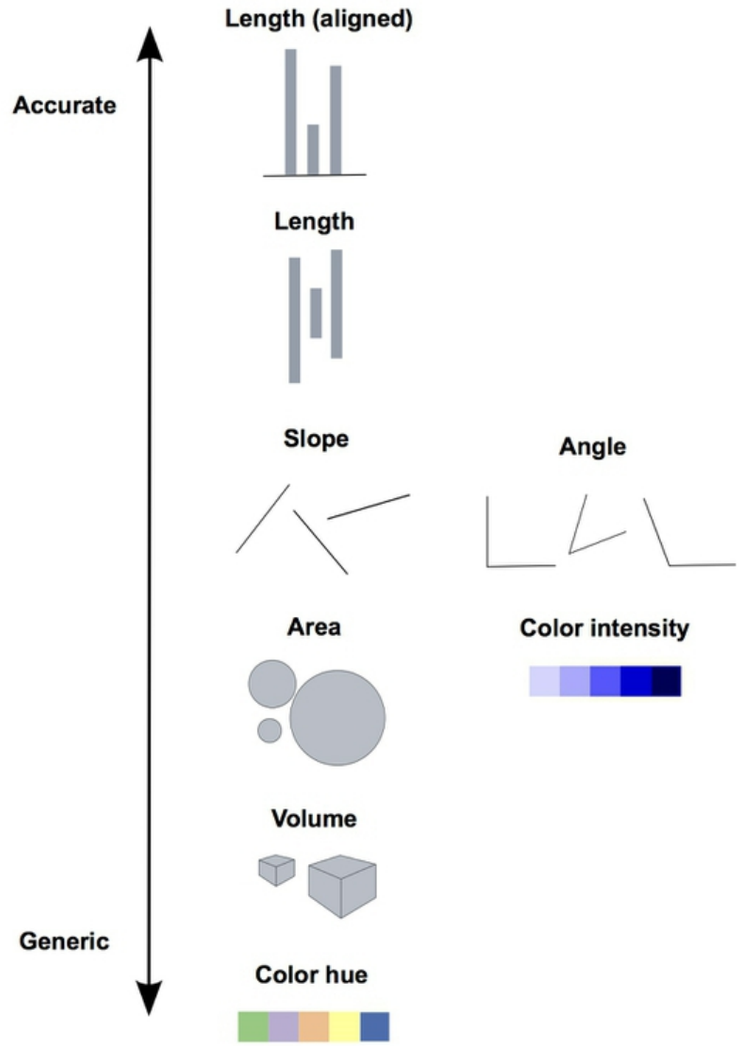


Diamonds

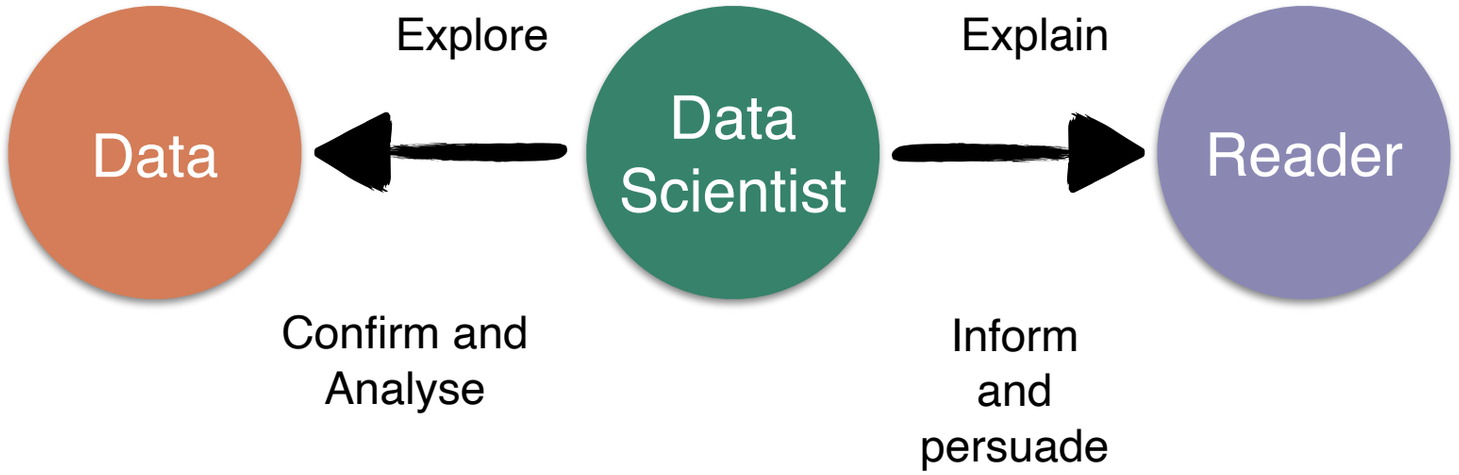

UNDERSTANDING THE FOUR C'S
CLARITY
International Grading Scale
loupe clean
VVS1
VVS2
VS1
VS2
SI1
SI2
P1 OR I1
P2 OR I2
P3 OR I3
COLOUR
International Grading Scale
D
Exceptional White +
E
Exceptional White
F
Rare White +
G
Rare White
H
White
I
Slightly Tinted White
J
Slightly Tinted White
K
Tinted White
L
Tinted White
M-Z
Tinted Colour
CARAT WEIGHT
Size Indication Scale
3.00 CARAT 9.4mm
2.50 CARAT 8.8mm
2.00 CARAT 8.2mm
1.75 CARAT 7.8mm
1.50 CARAT 7.4mm
1.25 CARAT 7mm
1.00 CARAT 6.5mm
0.50 CARAT 5.2mm
0.25 CARAT 4.1mm

Form follows function

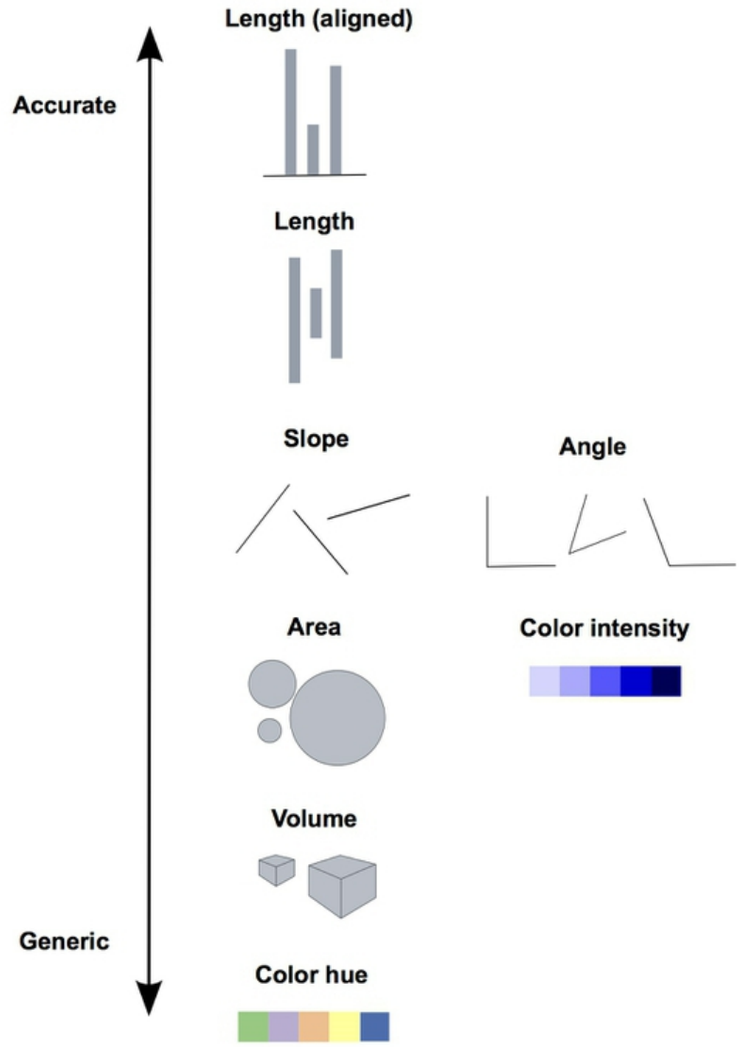




Form follows function

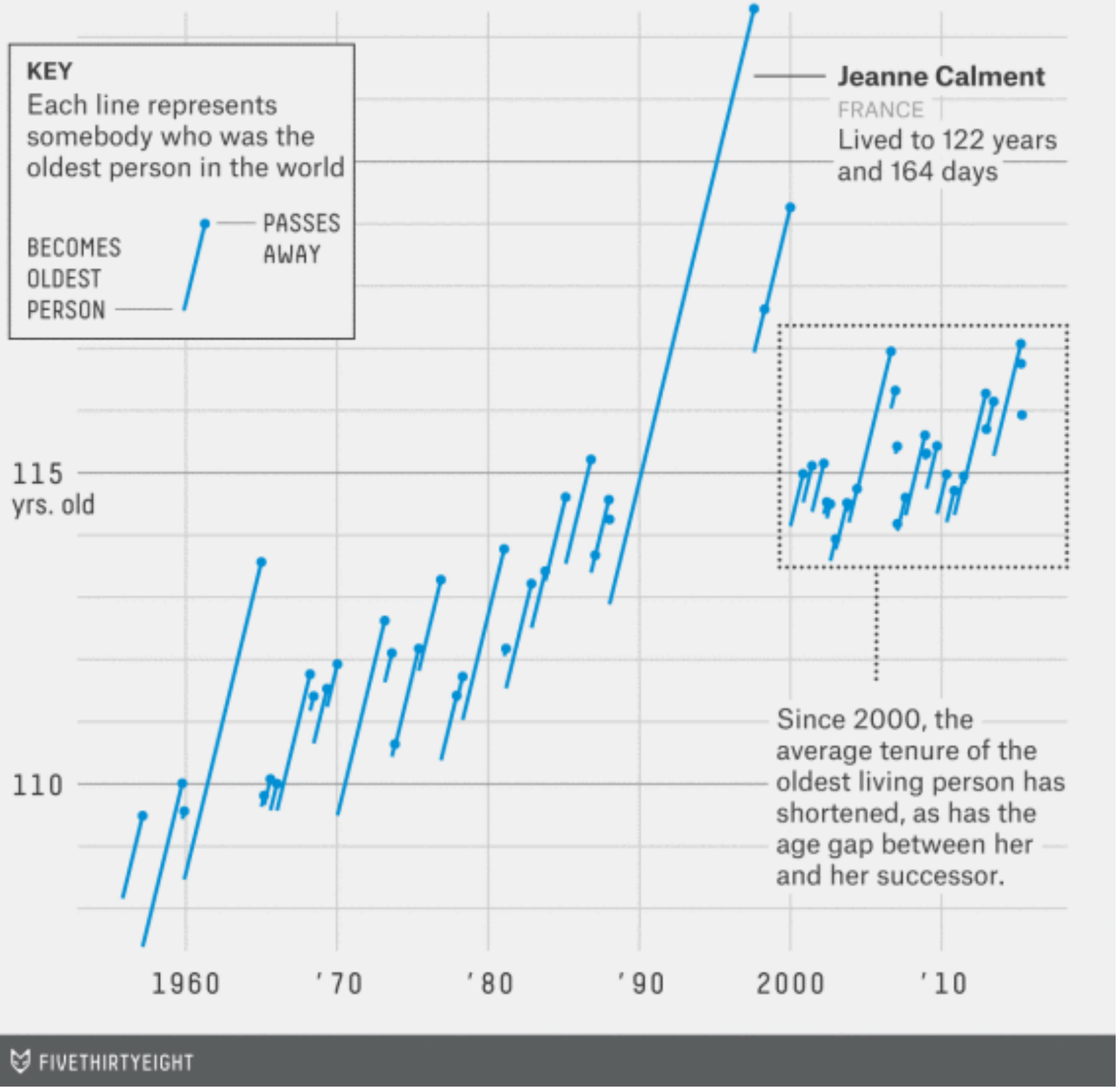


Think about the audience!



Some Data Vis. Examples from the Wider World

The Oldest Persons In The World

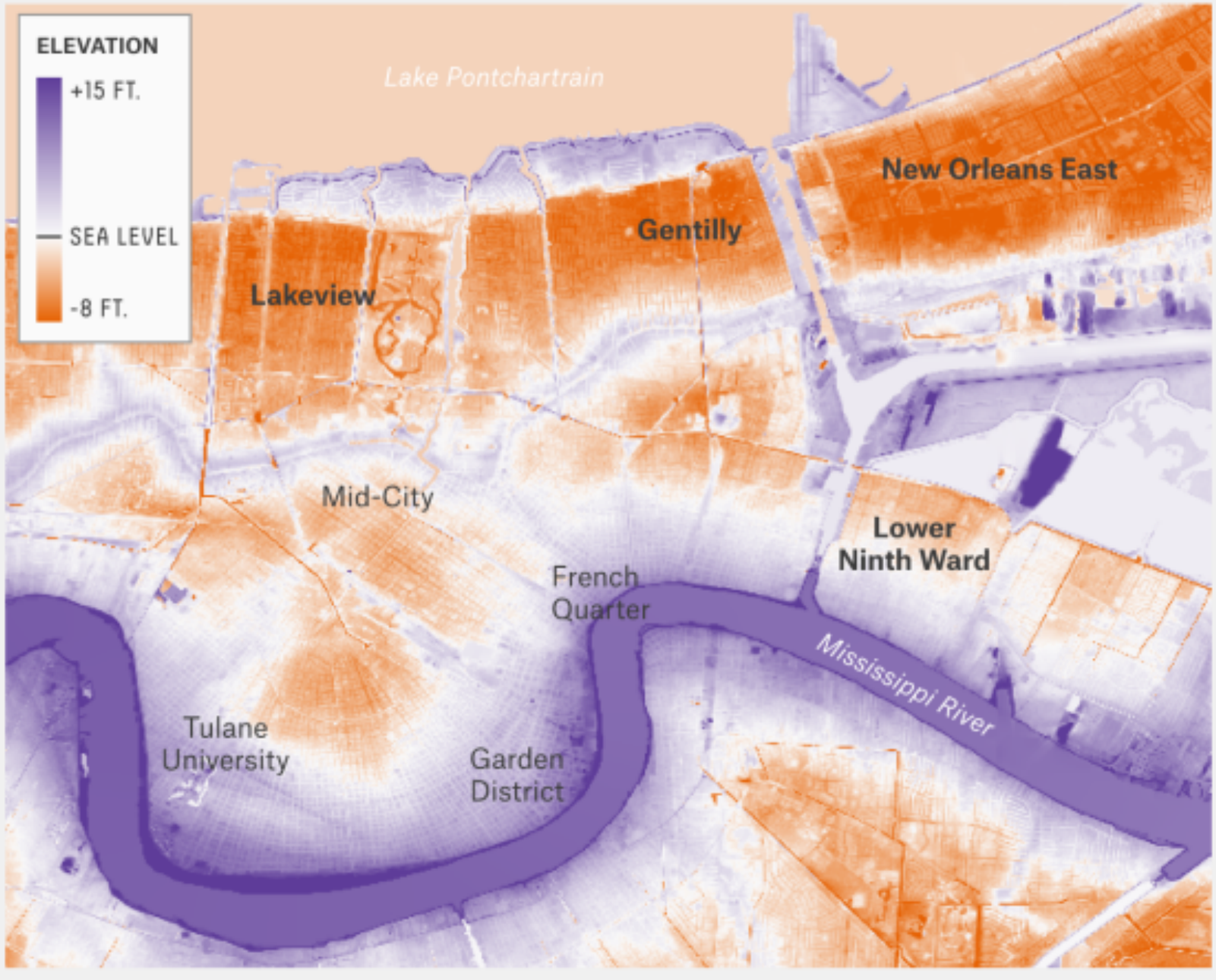


FIVETHIRTYEIGHT

https://fivethirtyeight.com

New Orleans, Under Water

Elevation in the city, orange areas below sea level



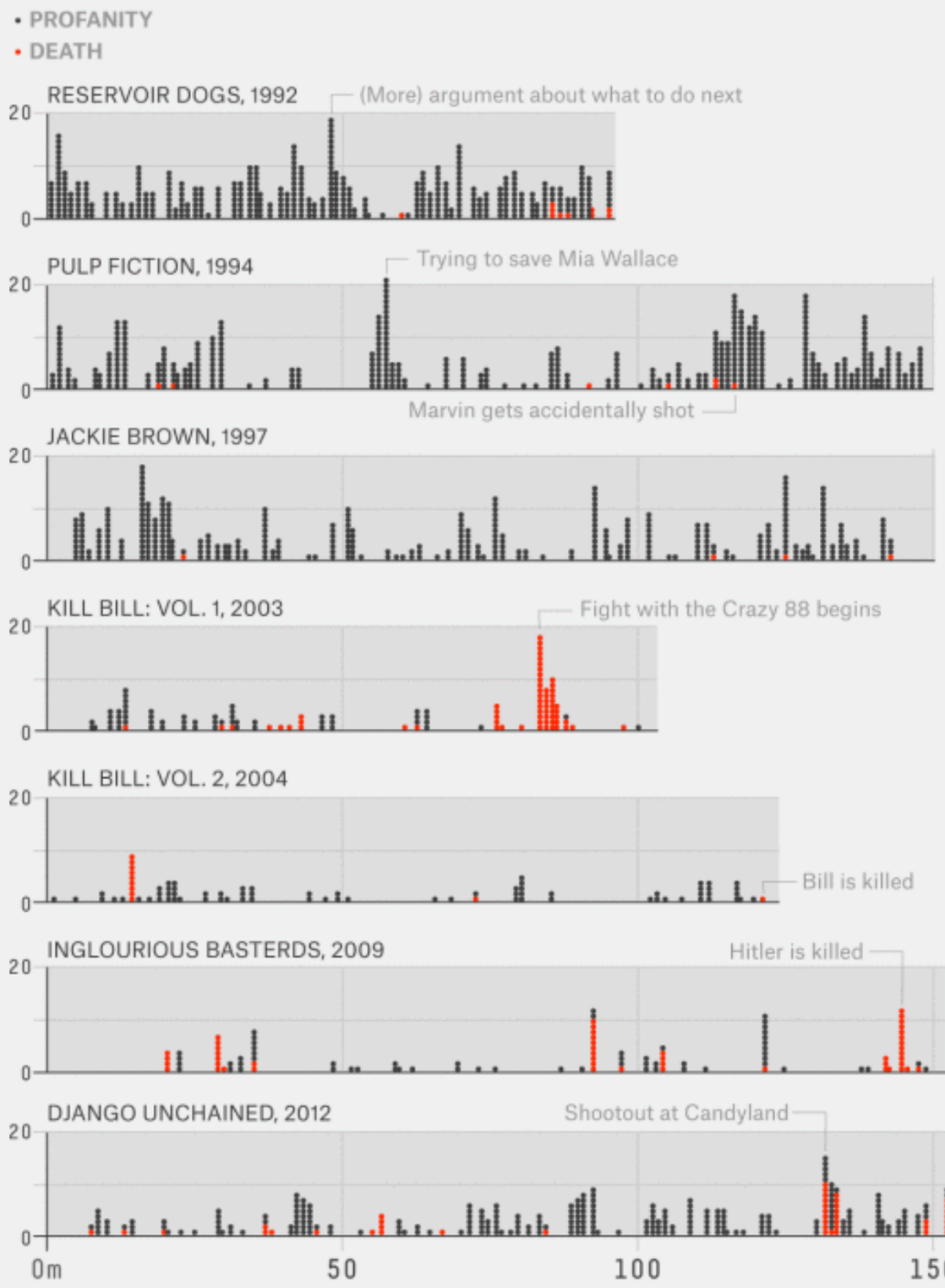
REUBEN FISCHER-BAUM

SOURCE: USGS

https://fivethirtyeight.com

The complete obscene guide to Tarantino

Time stamp of every instance of profanity and each death in feature films directed by Quentin Tarantino

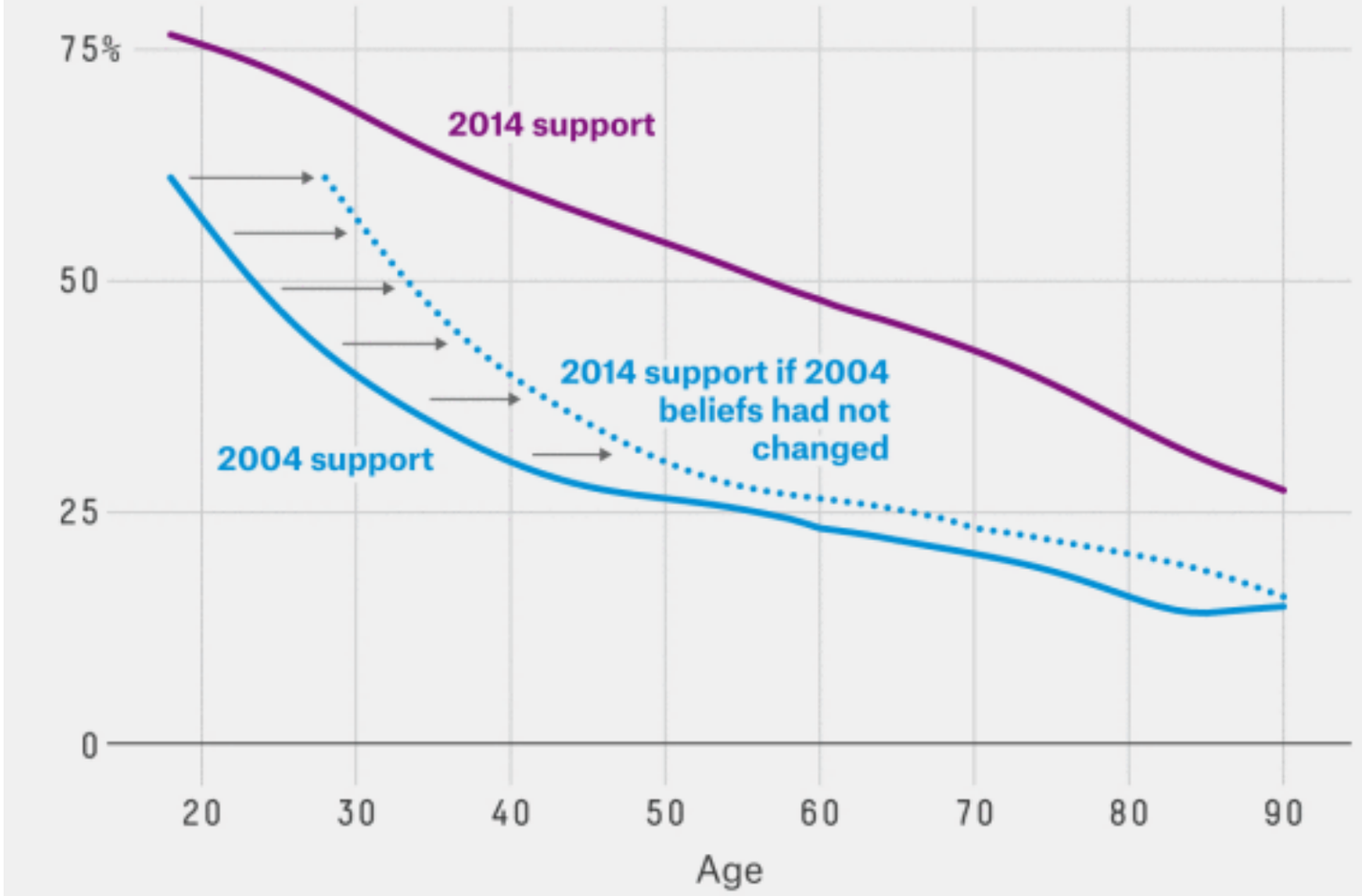


FIVETHIRTYEIGHT

https://fivethirtyeight.com

Minds Have Changed On Gay Marriage, Not Just Populations

Estimated support for gay marriage by age, 2004 and 2014

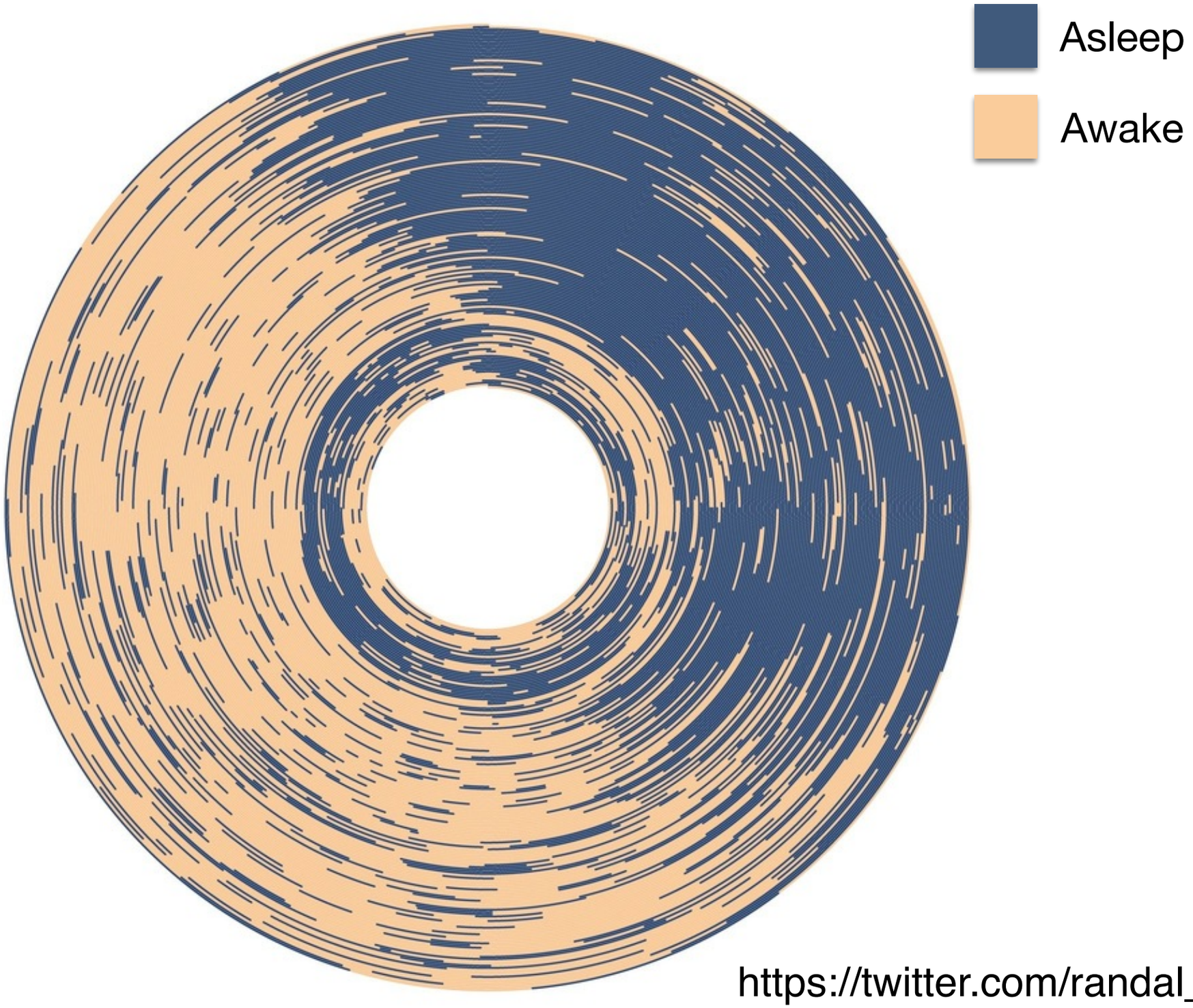


FIVETHIRTYEIGHT

SOURCE: GENERAL SOCIAL SURVEY

https://fivethirtyeight.com

- Sleeping patterns for babies first 4 months of her life.
- One continuous spiral starting on the inside when she was born.
- Each revolution representing a single day.
- Midnight at the top (24 hour clock)



https://twitter.com/randal_olson

Where to get help?

- ggplot docs are very good:
  http://docs.ggplot2.org/current/

- DataCamp online course:
  https://www.datacamp.com/courses/data-visualization-with-ggplot2-1

- R Cookbook - Graphics:
  http://www.cookbook-r.com/Graphs/

- Stackoverflow but beware!

Conclusion

- Telling visual stories with data:
  - Explorative and Explanatory
- A grammar of graphics underpins data visualisation
- Structure your data and your working to support data visualisation and reproducibility
- Keep your hypothesis and reader in mind

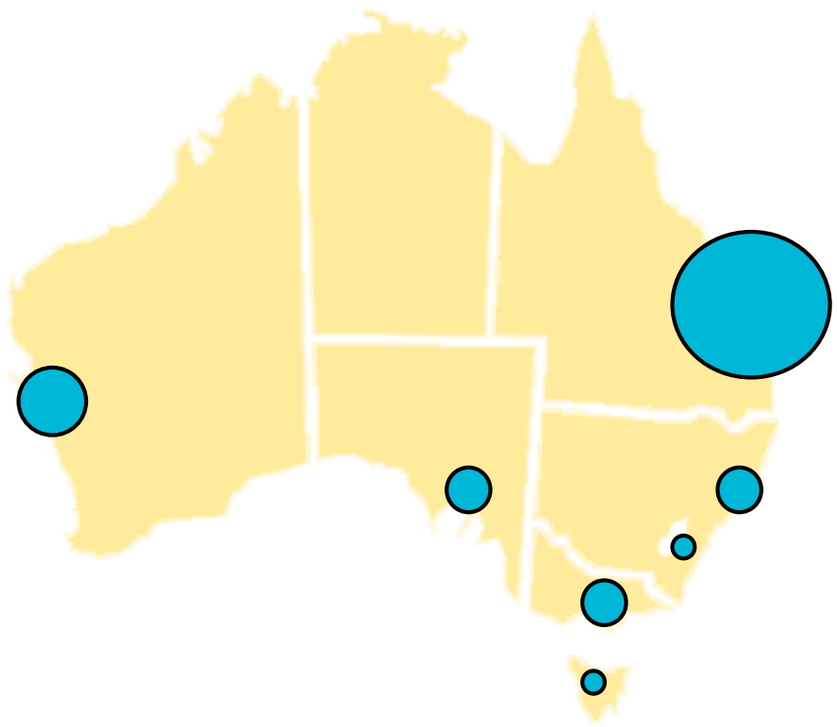
Australian e-Health Research Centre

- AeHRC is the leading national eHealth research group in Australia
- Currently 60-70 staff, students, visiting researchers

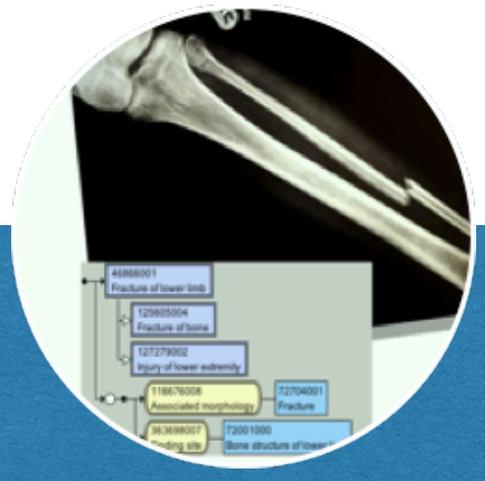
THE AUSTRALIAN E-HEALTH RESEARCH CENTRE

CSIRO

Queensland Health



Research Areas

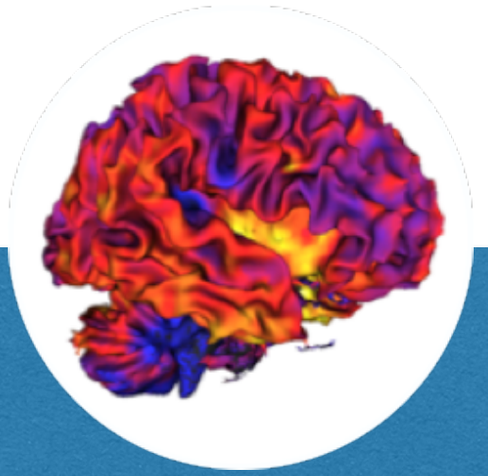


Health Data Semantics

- IR
- NLP
- Ontologies
- Semantic Web

Health Services

- Mobile/Tele Health
- Forecasting



Biomedical Informatics

- Medical Imaging
- Biostatistics

Come Work for Us!

- Undergraduate projects
- Vacation students
- Interns and casual employment
- PhD and Masters projects
- Research staff (PhD qualified)
- Software engineers

THE AUSTRALIAN E-HEALTH RESEARCH CENTRE

CSIRO

aehrc.com

My details

Dr. Bevan Koopman

Australian e-Health Research Centre
CSIRO

www: http://koopman.id.au
@bevan_koopman

THE AUSTRALIAN E-HEALTH RESEARCH CENTRE

CSIRO