Strategic Data Science (SDS)

Data Visualization

Karl Ho
School of Economic, Political and Policy Sciences
University of Texas at Dallas

What is data visualization?

This is a course training students in:

What is data visualization?

This is a course training students in:

- Data Literacy
  - Understand data theory
  - Manage data
  - Analyze data

What is data visualization?

This is a course training students in:

- Data Literacy
  - Understand data theory
  - Manage data
  - Analyze data
- Data Skills
  - Programming
  - Tools

What is data visualization?

Objectives:

What is data visualization?

Objectives:

1. Understand data theory

What is data visualization?

Objectives:

1. Understand data theory
2. Be familiar with principles behind effective data visualization

What is data visualization?

Objectives:

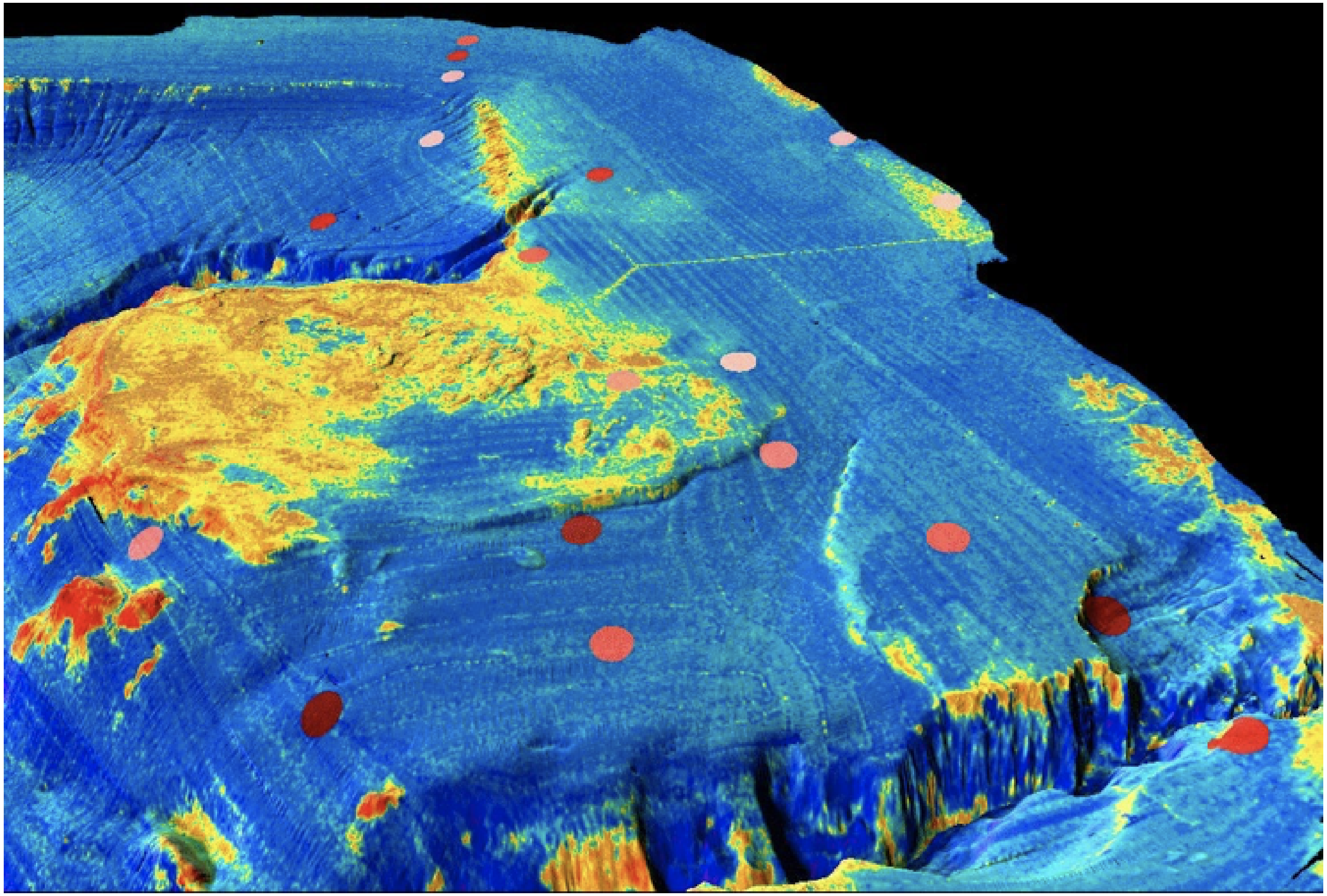
1. Understand data theory
2. Be familiar with principles behind effective data visualization
3. Communicate message using visual form of data

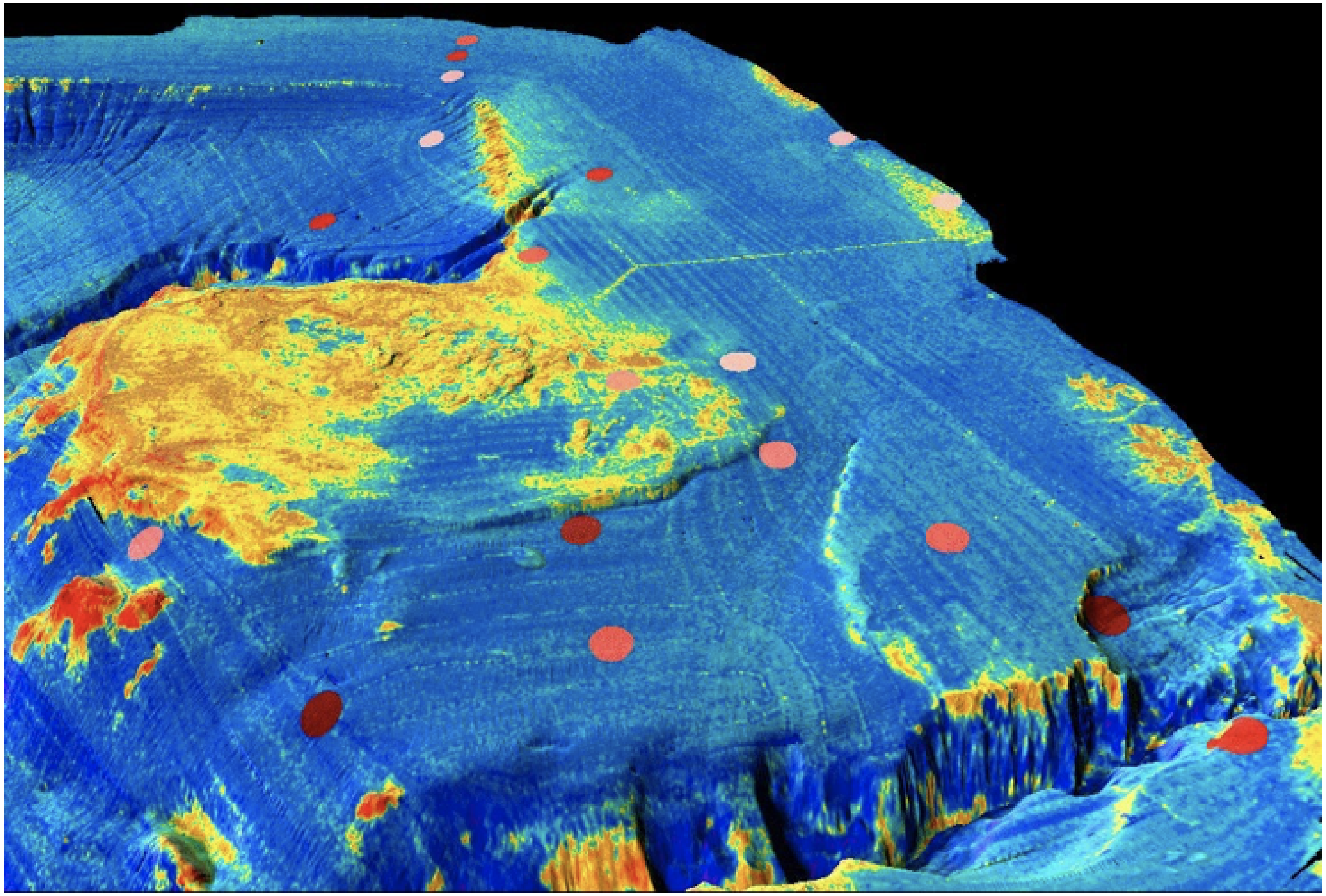
What is data visualization?

Objectives:

1. Understand data theory
2. Be familiar with principles behind effective data visualization
3. Communicate message using visual form of data
4. Conduct independent research focusing on visual data analytics

How much information does this picture present?





Multibeam sonar backscatter data draped on bathymetry off Santa Monica Calif. Yellow is high backscatter. Santa Monica sewer pipe and diffuser is visible in upper part of image (y-shaped feature). Red-brown dots represent color-coded fish abundance as determined from trawl data. Source: https://tinyurl.com/ydhqtr8f

What do we learn from the image?

What do we learn from the image?

- Visualization provides an ability to comprehend huge amounts of data. The important information from more than a million measurements is immediately available.

What do we learn from the image?

What do we learn from the image?

- Visualization often enables problems with the data to become immediately apparent. A visualization commonly reveals things not only about the data itself but also about the way it is collected. With an appropriate visualization, errors and artifacts in the data often jump out at you. For this reason, visualizations can be invaluable in quality control.

What do we learn from the image?

What do we learn from the image?

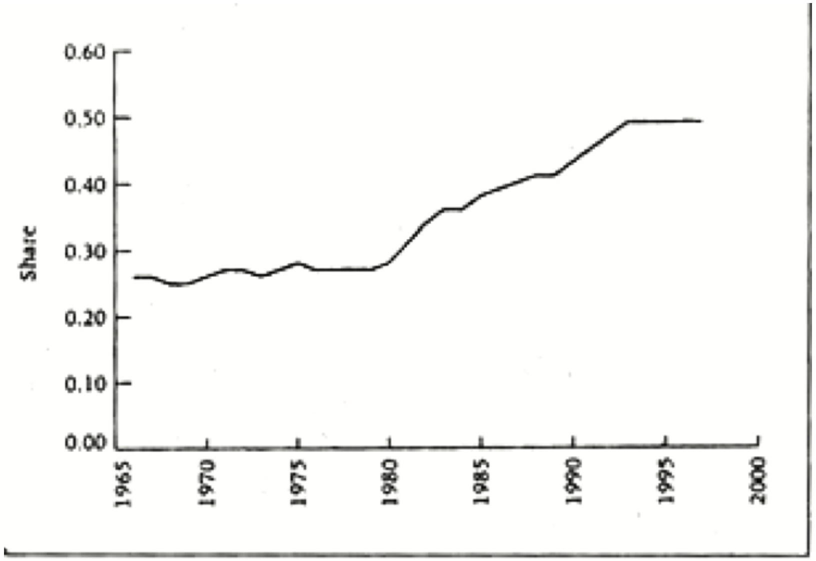
- Visualization facilitates understanding of both large-scale and small-scale features of the data. It can be especially valuable in allowing the perception of patterns linking local features.

What do we learn from the image?

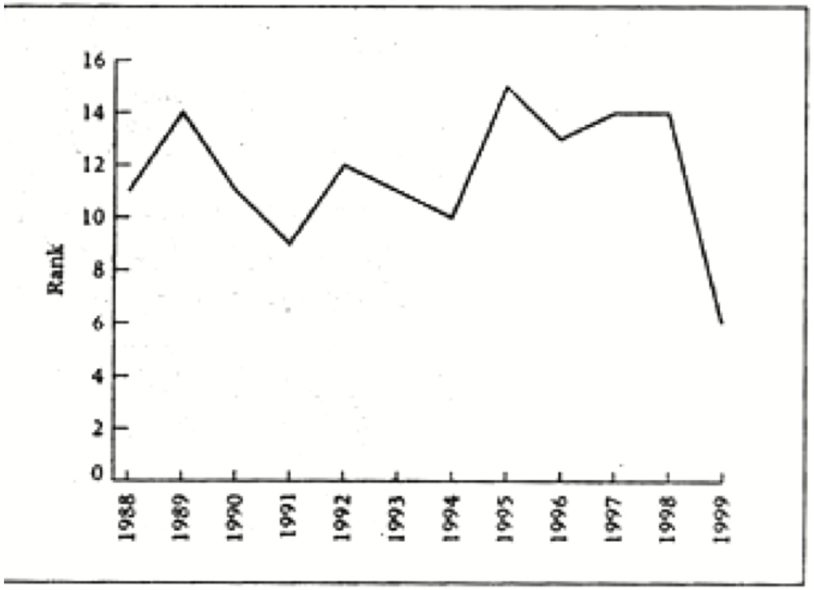
What do we learn from the image?

- Visualization facilitates hypothesis formation, inviting further inquiries into building a theory

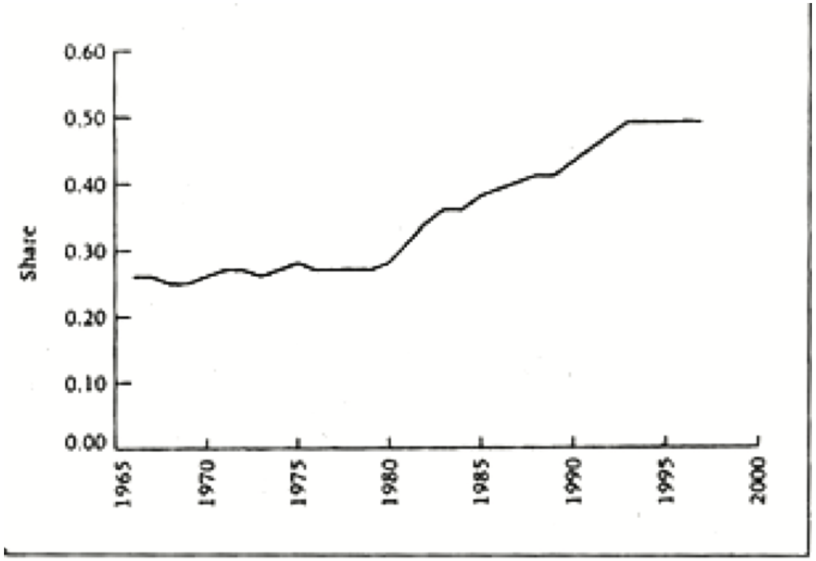
(Colin Ware 2012, Ch. 1)



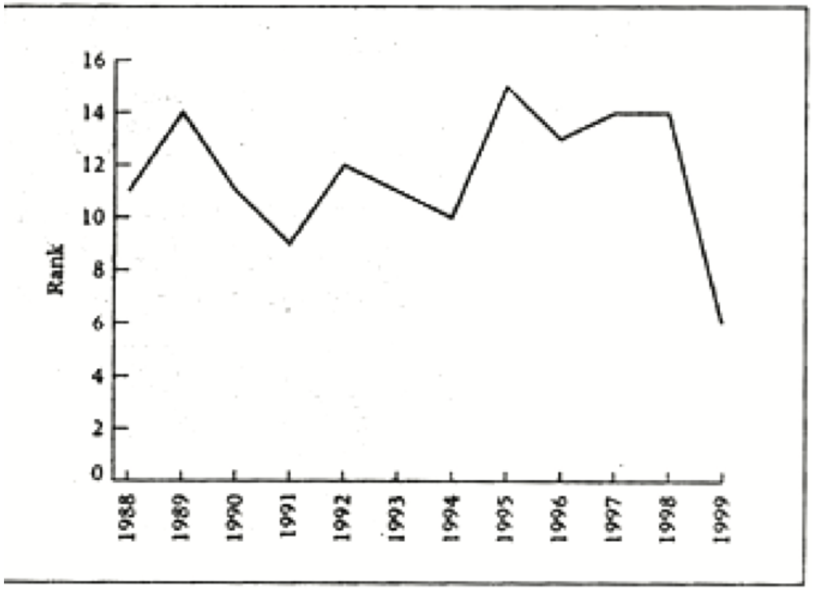
By the Numbers: Over 35 years, Cornell's tuition has taken an increasingly larger share of its median student family income.



Pecking Order: Over 12 years, Cornell's ranking in *US News & World Report* has risen and fallen erratically.

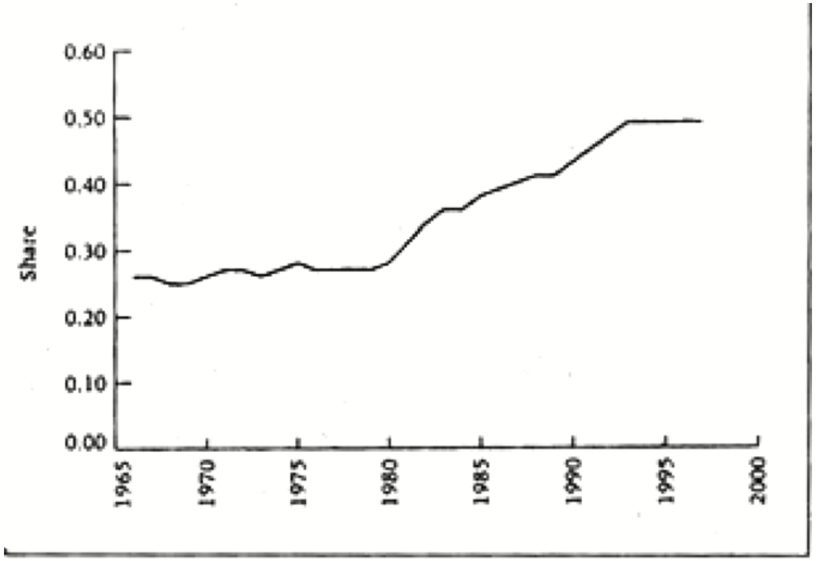


BY THE NUMBERS: OVER 35 YEARS, CORNELL'S TUITION HAS TAKEN AN INCREASINGLY LARGER SHARE OF ITS MEDIAN STUDENT FAMILY INCOME.

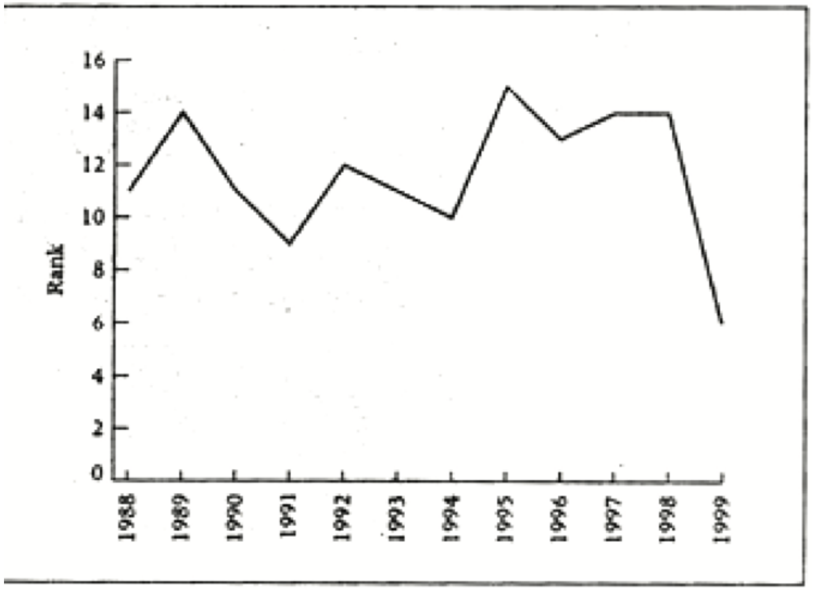


PECKING ORDER: OVER 12 YEARS, CORNELL'S RANKING IN *US NEWS & WORLD REPORT* HAS RISEN AND FALLEN ERRATICALLY.

Source: Chris Adolph, also Johnson, R.R. and Kuby, P.J., 2011. *Elementary statistics*. Cengage Learning.



By the Numbers: Over 35 years, Cornell's tuition has taken an increasingly larger share of its median student family income.

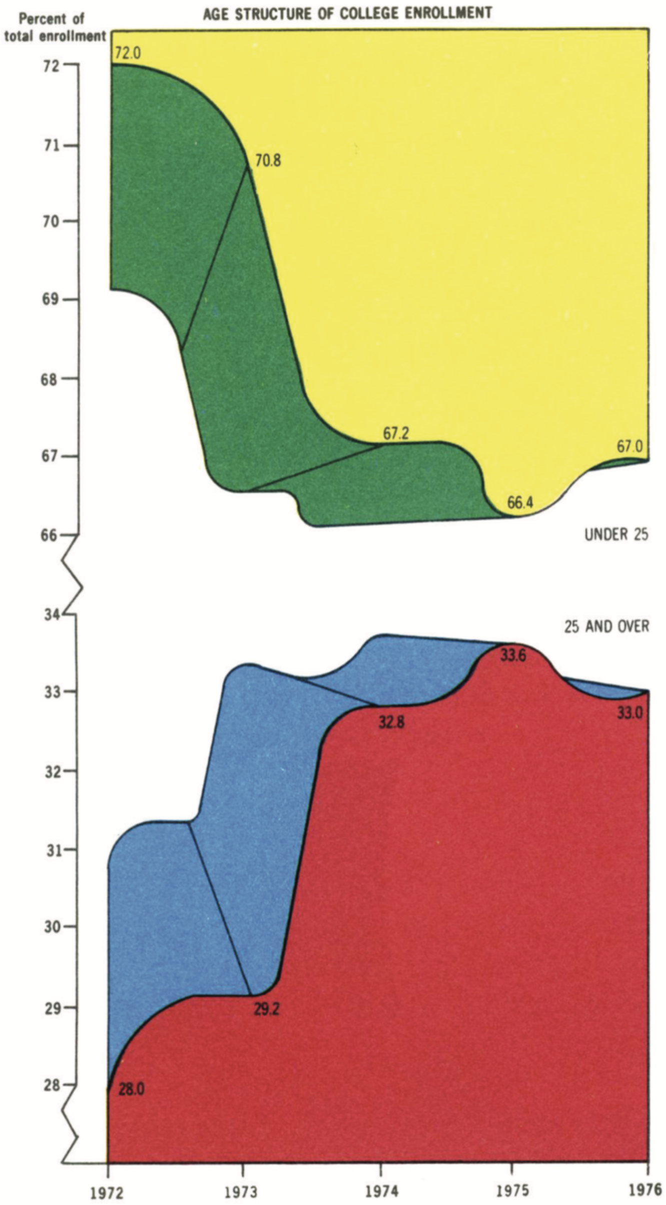


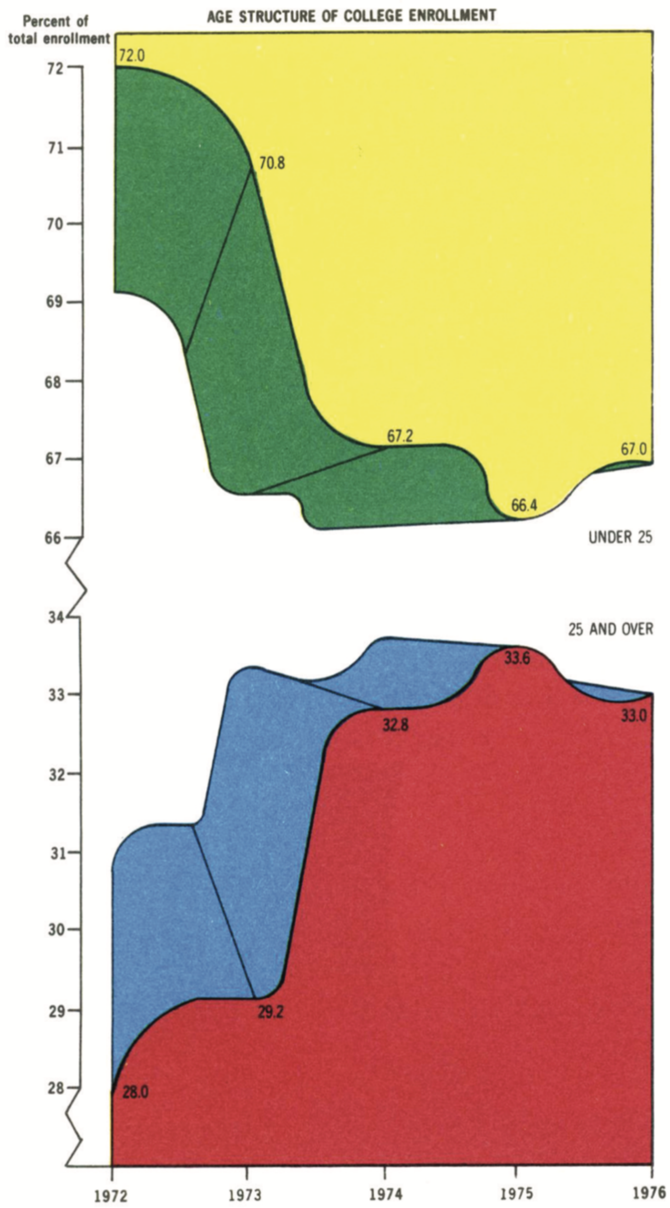
Pecking Order: Over 12 years, Cornell's ranking in *US News & World Report* has risen and fallen erratically.

Messages:

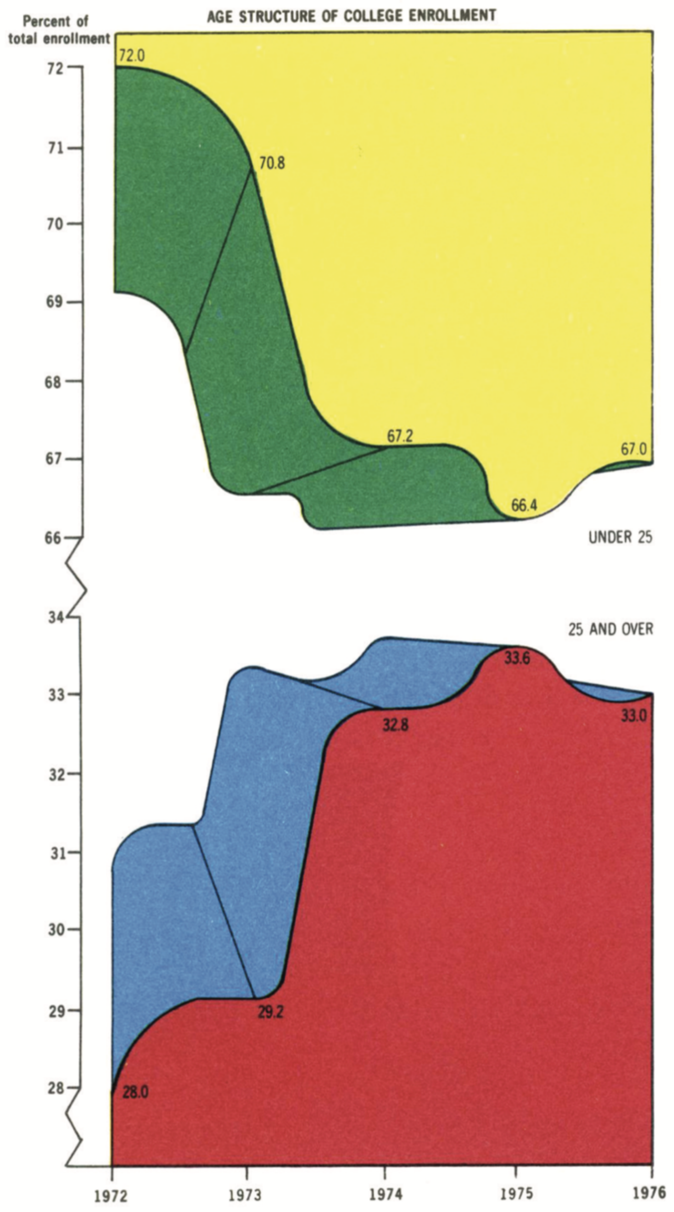
- Gradual rise?
- Abrupt Drop of Ranking?

Source: Chris Adolph, also Johnson, R.R. and Kuby, P.J., 2011. *Elementary statistics*. Cengage Learning.


AGE STRUCTURE OF COLLEGE ENROLLMENT
Percent of total enrollment
72
71
70
69
68
67
66
34
33
32
31
30
29
28
72.0
70.8
67.2
67.0
66.4
UNDER 25
25 AND OVER
33.6
32.8
33.0
29.2
28.0
1972
1973
1974
1975
1976



Source: Edward R. Tufte. 2001. The Visual Display of Quantitative Information. Graphics Press. 2nd ed.

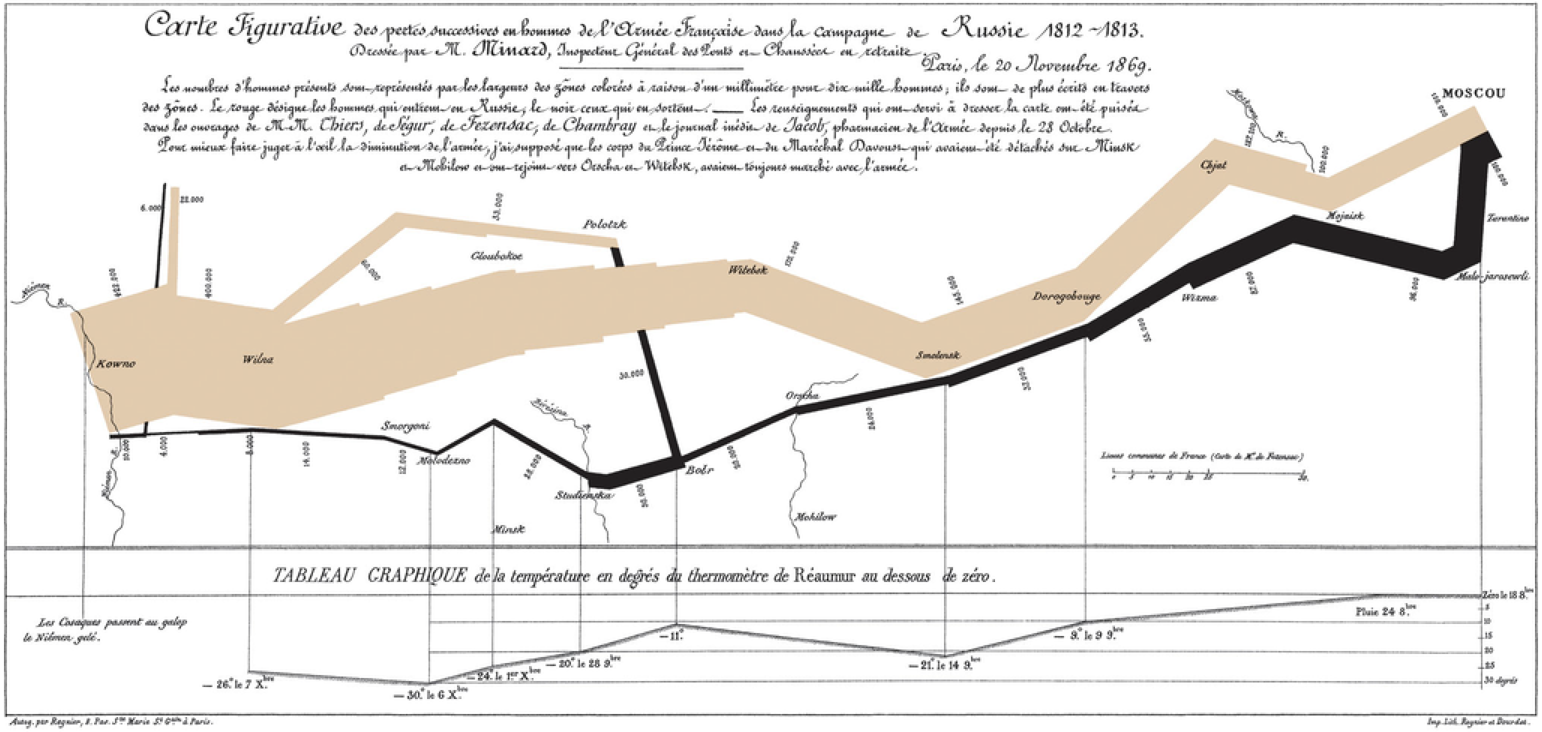


Message:

- Age structure of college enrollment
- How much data are presented in multiple colors?

Source: Edward R. Tufte. 2001. The Visual Display of Quantitative Information. Graphics Press. 2nd ed.

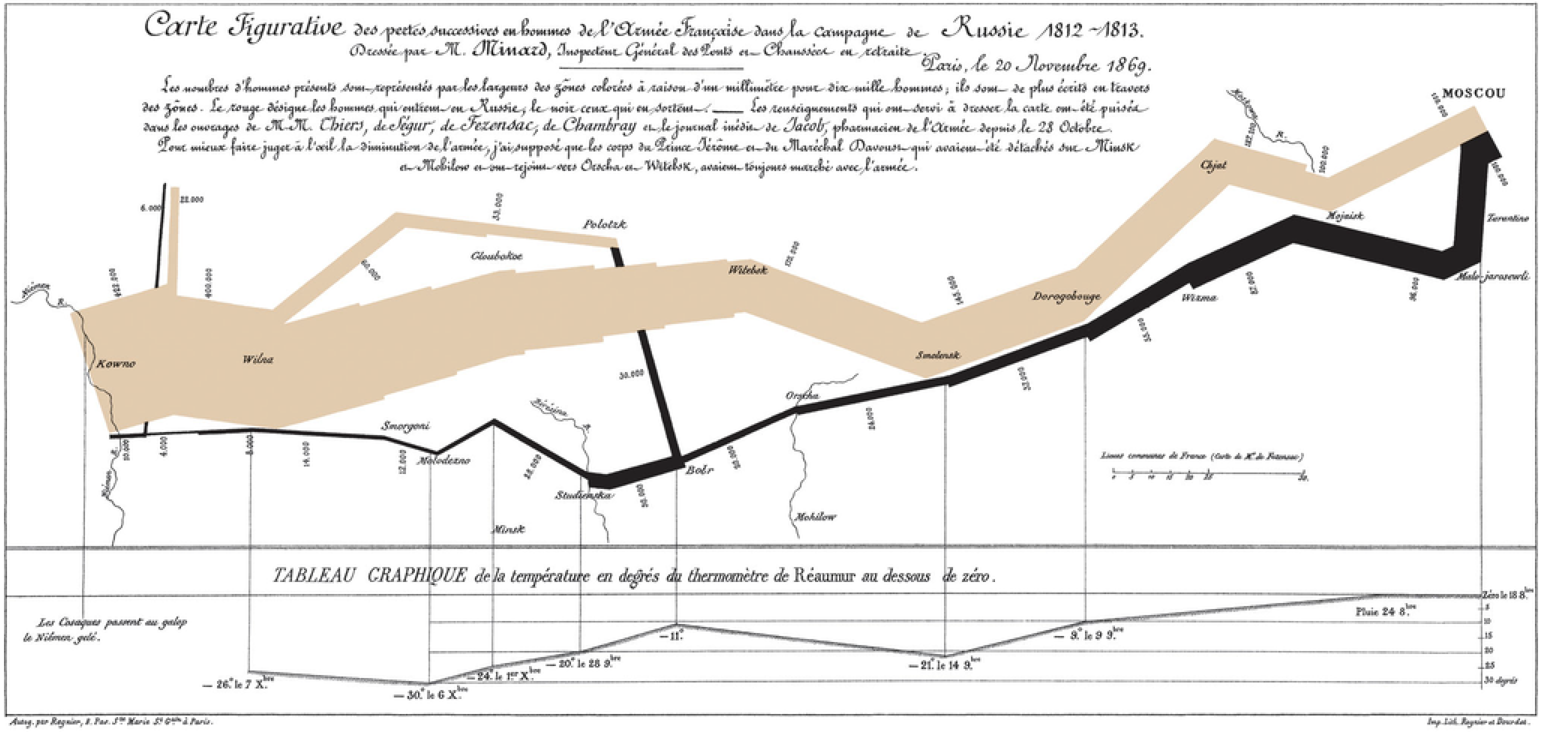
One of the best data visualizations in history



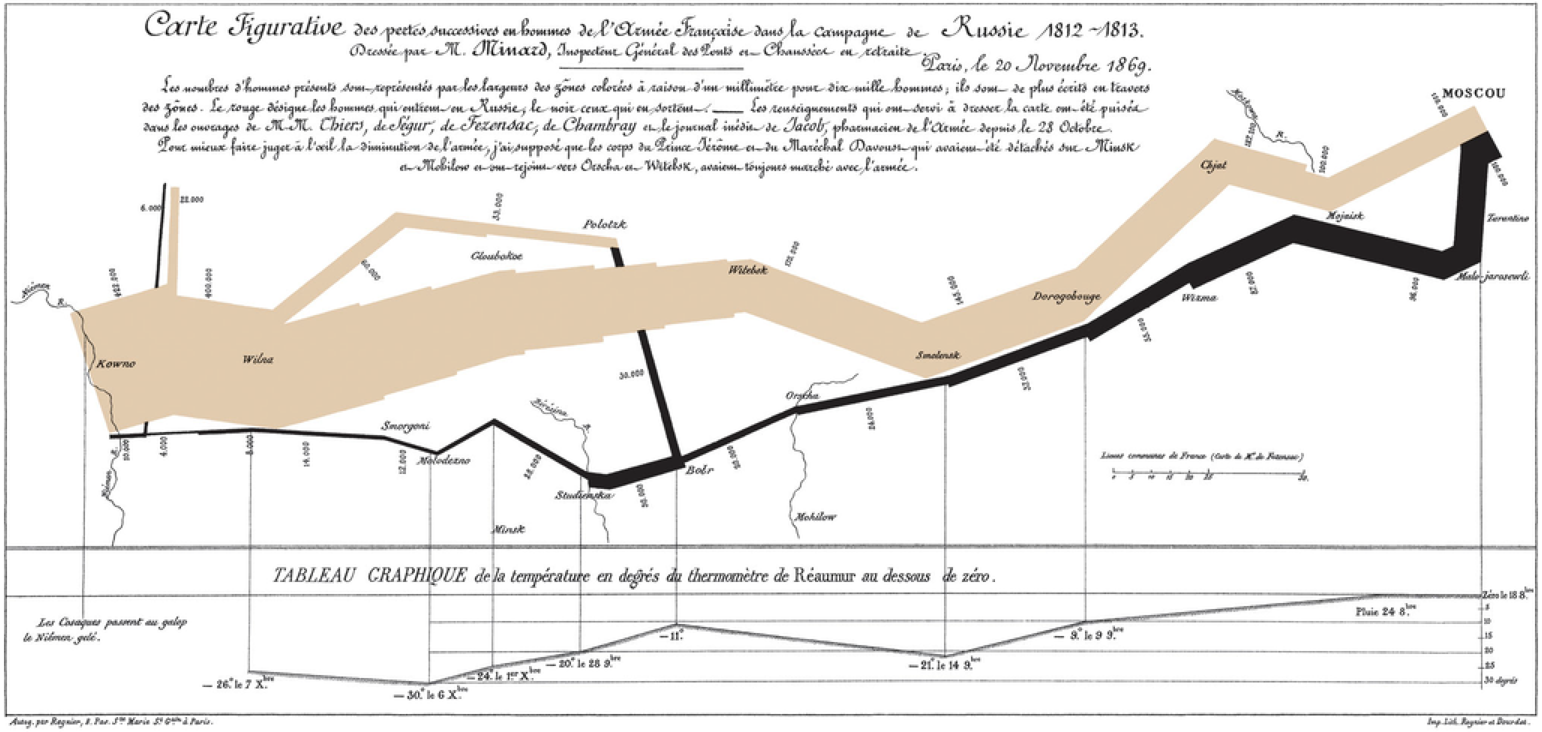
Charles Joseph Minard, in mapping Napoleon's march on Moscow

Source: https://en.wikipedia.org/wiki/Charles_Joseph_Minard

One of the best data visualizations in history

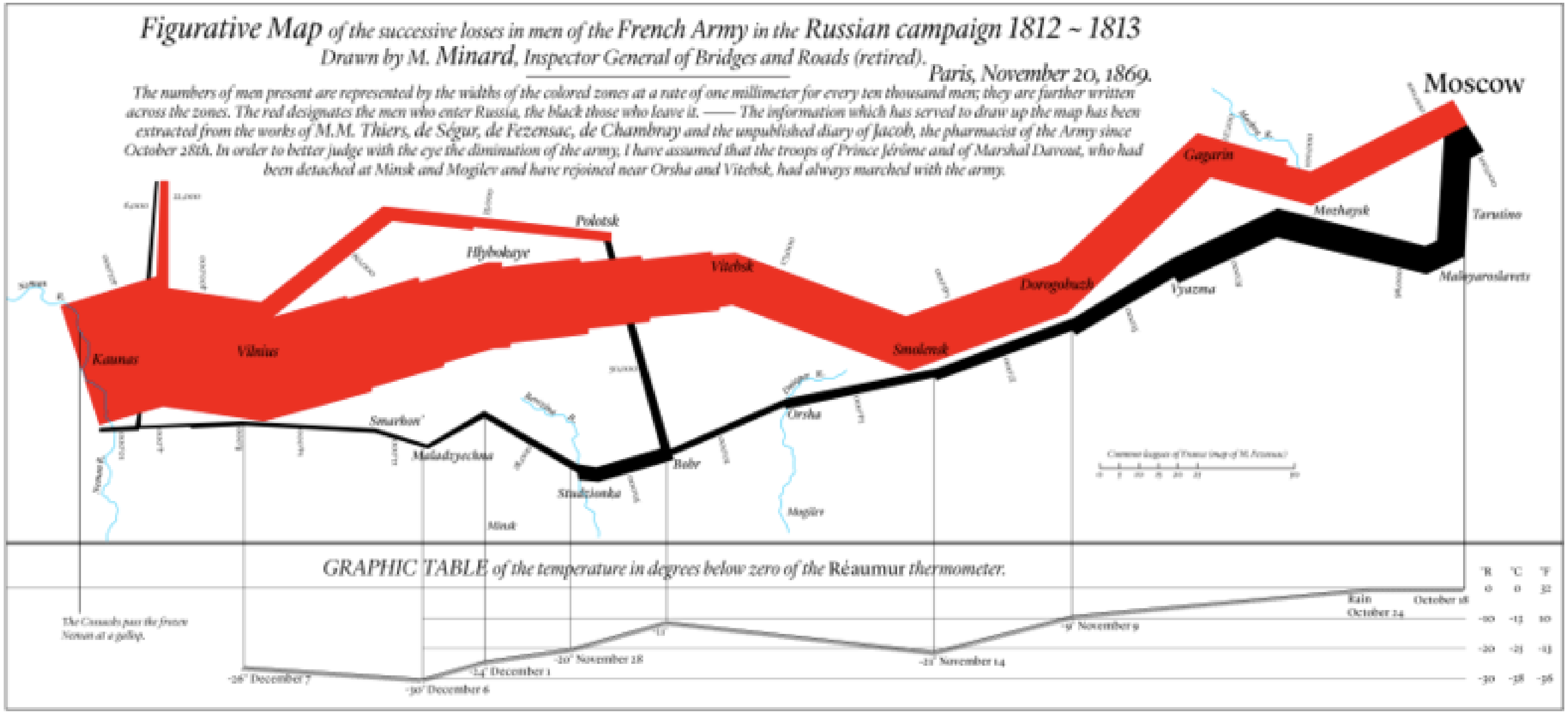


One of the best data visualizations in history



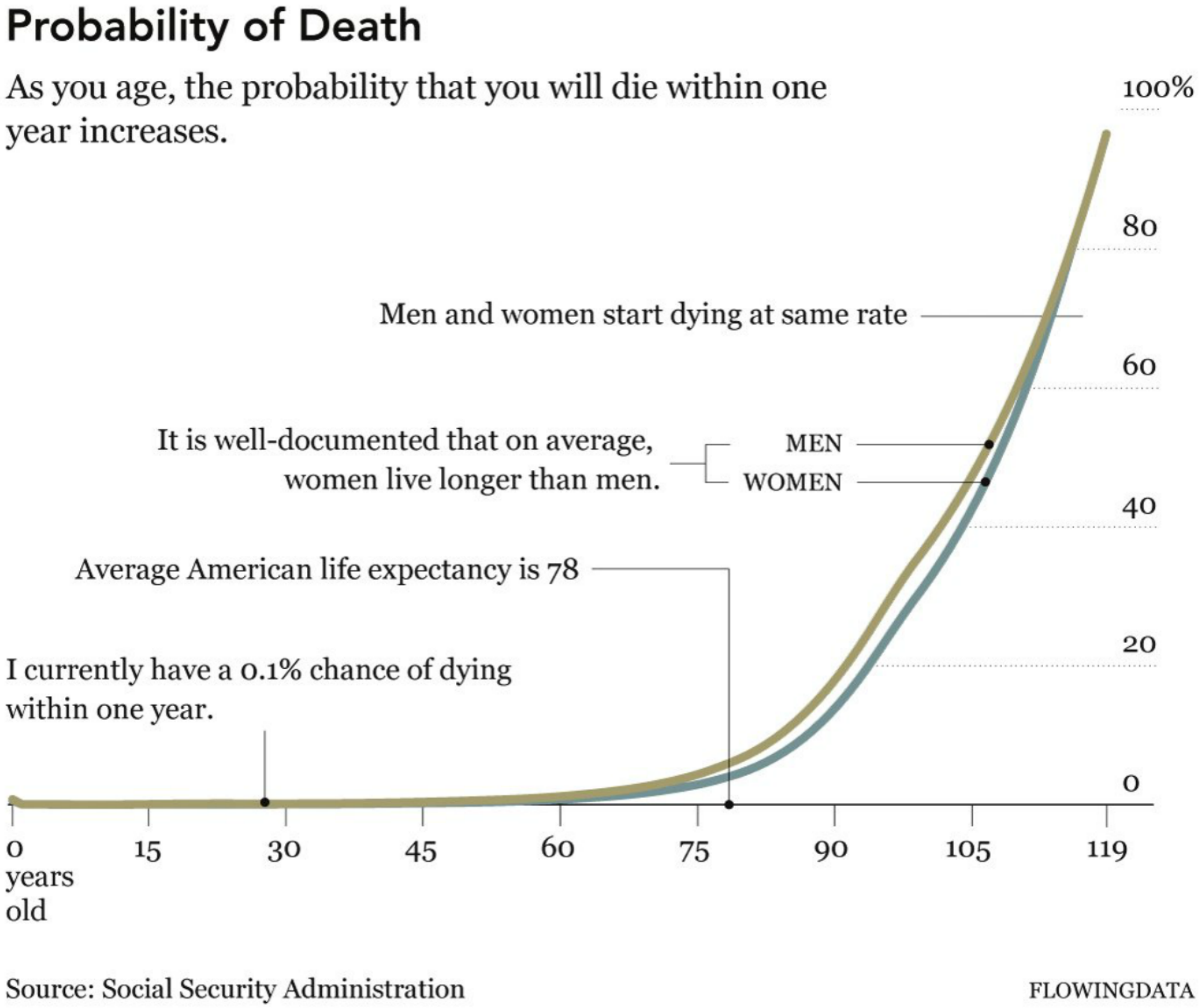
How much information?

1. Latitude of army & features (Y-coordinate) .
2. Longitude of army & features (X-coordinate)
3. Size of army (width of line, numerals) .
4. Advance vs. Retreat color of line
5. Division of army splitting of line
6. Temperature linked lineplot
7. Time linked lineplot



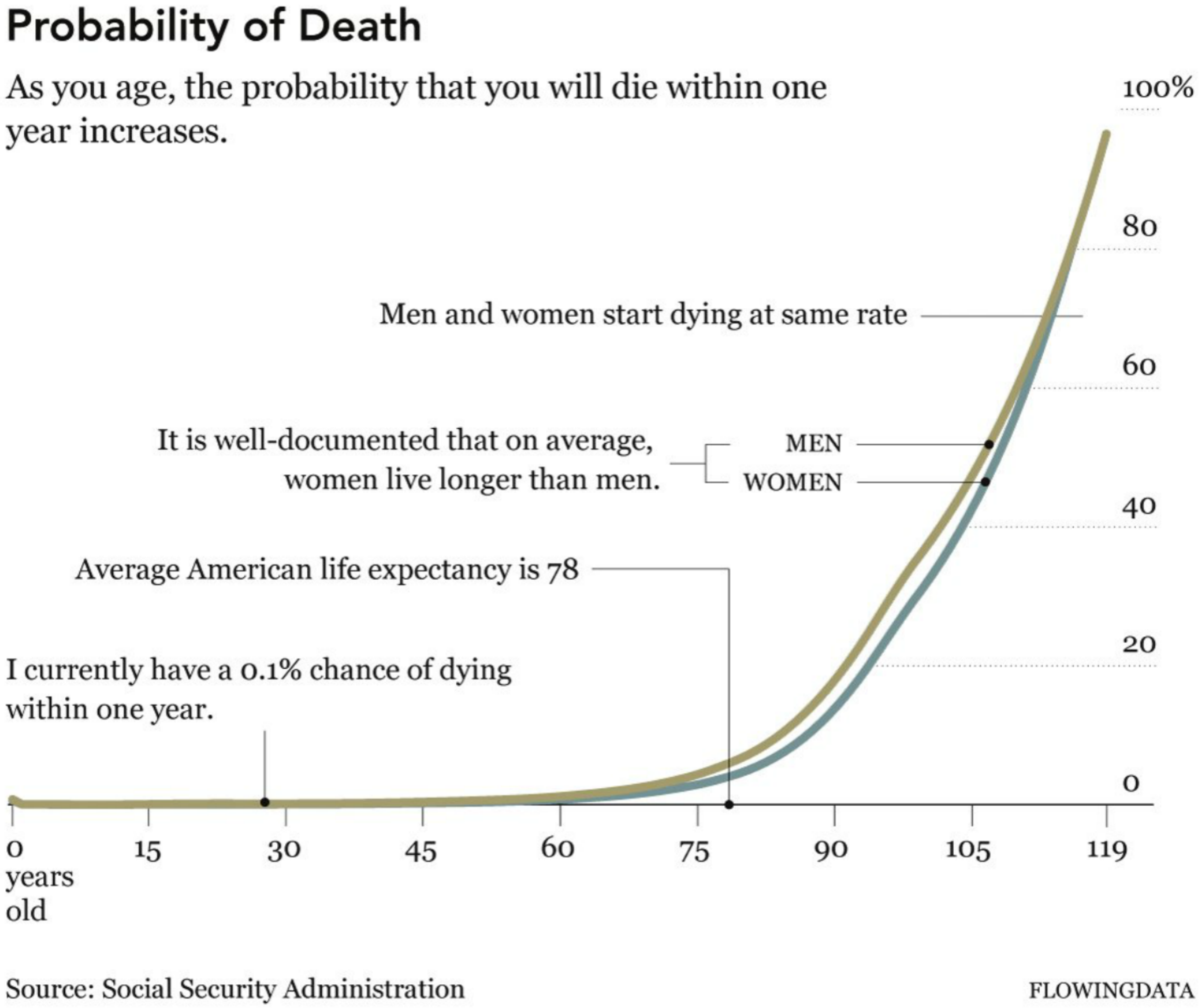
Source: https://en.wikipedia.org/wiki/Charles_Joseph_Minard

Data Story:



Probability of Death
As you age, the probability that you will die within one year increases.
100%
80
60
40
20
0
Men and women start dying at same rate
It is well-documented that on average, women live longer than men.
MEN
WOMEN
Average American life expectancy is 78
I currently have a 0.1% chance of dying within one year.
0 years old
15
30
45
60
75
90
105
119
Source: Social Security Administration
FLOWINGDATA

Data Story:



Source: Yau 2011

Numbers

Numbers

1. Numbers vs. ?

Numbers

1. Numbers vs. ?
2. Zeros and ones

Numbers

1. Numbers vs. ?
2. Zeros and ones
3. Integers (0, 1, 2, 100) vs. Non-integers (2.5, 3.1416)

Numbers

1. Numbers vs. ?
2. Zeros and ones
3. Integers (0, 1, 2, 100) vs. Non-integers (2.5, 3.1416)
4. Positive and negative numbers

Quantitative vs. Qualitative Data

Quantitative vs. Qualitative Data

1. Numbers vs. Labels

Quantitative vs. Qualitative Data

1. Numbers vs. Labels
2. Quantity vs. Quality

Quantitative vs. Qualitative Data

1. Numbers vs. Labels
2. Quantity vs. Quality
3. Ordinal, Interval, Ratio vs. Nominal

Quantitative vs. Qualitative Data

1. Numbers vs. Labels
2. Quantity vs. Quality
3. Ordinal, Interval, Ratio vs. Nominal
4. e.g. Yes/No--> Qualitative

Quantitative vs. Qualitative Data

1. Numbers vs. Labels
2. Quantity vs. Quality
3. Ordinal, Interval, Ratio vs. Nominal
4. e.g. Yes/No--> Qualitative
5. e.g. How much--> Quantitative

Quantitative vs. Qualitative Data

Quantitative vs. Qualitative Data

1. Higher quantity means higher quality?

Quantitative vs. Qualitative Data

1. Higher quantity means higher quality?
2. Higher quality leads to higher quantity?

Basics of data organization

Basics of data organization

1. Variables and observations

Basics of data organization

1. Variables and observations
    1. Alternative terms: Fields and cases

Basics of data organization

1. Variables and observations
    1. Alternative terms: Fields and cases
2. Rows for observations , columns for variables

Basics of data organization

1. Variables and observations
    1. Alternative terms: Fields and cases
2. Rows for observations , columns for variables
3. Names and labels

Basics of data organization

1. Variables and observations
    1. Alternative terms: Fields and cases
2. Rows for observations , columns for variables
3. Names and labels
4. Table vs. query

What to visualize in data?

What to visualize in data?

1. Data Generating Process

What to visualize in data?

1. Data Generating Process
2. Property

What to visualize in data?

1. Data Generating Process
2. Property
3. Distribution

What to visualize in data?

1. Data Generating Process
2. Property
3. Distribution
4. Pattern

What to visualize in data?

1. Data Generating Process
2. Property
3. Distribution
4. Pattern
5. Differences

What to visualize in data?

1. Data Generating Process
2. Property
3. Distribution
4. Pattern
5. Differences
6. Relationship

Time series data

Time series data

1. Nature

Time series data

1. Nature
    1. Temporal dependency: non-stationarity autocorrelation

Time series data

1. Nature
    1. Temporal dependency: non-stationarity autocorrelation
    2. Periodicity: seasonality, cycle

Time series data

1. Nature
   1. Temporal dependency: non-stationarity autocorrelation
   2. Periodicity: seasonality, cycle
2. Zeros -> events?

Time series data

1. Nature
   1. Temporal dependency: non-stationarity autocorrelation
   2. Periodicity: seasonality, cycle
2. Zeros -> events?
3. Scale linearity

Event count data

Event count data

1. Nature

Event count data

1. Nature
   1. Distribution

Event count data

1. Nature
    1. Distribution
    2. Bounds

Event count data

1. Nature
   1. Distribution
   2. Bounds
      1. No upper bounds

Event count data

1. Nature
    1. Distribution
    2. Bounds
        1. No upper bounds
        2. One lower bound: zero

Event count data

1. Nature
    1. Distribution
    2. Bounds
        1. No upper bounds
        2. One lower bound: zero
    3. Zeros

Event count data

1. Nature
   1. Distribution
   2. Bounds
      1. No upper bounds
      2. One lower bound: zero
   3. Zeros
2. Continuous vs. discrete

Event count data

1. Nature
   1. Distribution
   2. Bounds
      1. No upper bounds
      2. One lower bound: zero
   3. Zeros
2. Continuous vs. discrete
3. Intervals vs. duration

Bertram M. Gross (1986)

"the world or my part of it is seen as an ongoing stream of events in time . . . Facts and process are separated into discrete elements only by human analysis . . . Change-whether rapid or slow, hidden or open-is continuous."

Anscombe example (1973)

```
> anscombe
   x1 x2 x3 x4    y1   y2    y3    y4
1  10 10 10  8  8.04 9.14  7.46  6.58
2   8  8  8  8  6.95 8.14  6.77  5.76
3  13 13 13  8  7.58 8.74 12.74  7.71
4   9  9  9  8  8.81 8.77  7.11  8.84
5  11 11 11  8  8.33 9.26  7.81  8.47
6  14 14 14  8  9.96 8.10  8.84  7.04
7   6  6  6  8  7.24 6.13  6.08  5.25
8   4  4  4 19  4.26 3.10  5.39 12.50
9  12 12 12  8 10.84 9.13  8.15  5.56
10  7  7  7  8  4.82 7.26  6.42  7.91
11  5  5  5  8  5.68 4.74  5.73  6.89
```

Anscombe example (1973)

```
> summary(anscombe)
       x1             x2             x3             x4           y1               y2       
 Min.   : 4.0   Min.   : 4.0   Min.   : 4.0   Min.   : 8   Min.   : 4.260   Min.   :3.100  
 1st Qu.: 6.5   1st Qu.: 6.5   1st Qu.: 6.5   1st Qu.: 8   1st Qu.: 6.315   1st Qu.:6.695  
 Median : 9.0   Median : 9.0   Median : 9.0   Median : 8   Median : 7.580   Median :8.140  
 Mean   : 9.0   Mean   : 9.0   Mean   : 9.0   Mean   : 9   Mean   : 7.501   Mean   :7.501  
 3rd Qu.:11.5   3rd Qu.:11.5   3rd Qu.:11.5   3rd Qu.: 8   3rd Qu.: 8.570   3rd Qu.:8.950  
 Max.   :14.0   Max.   :14.0   Max.   :14.0   Max.   :19   Max.   :10.840   Max.   :9.260  
       y3              y4        
 Min.   : 5.39   Min.   : 5.250  
 1st Qu.: 6.25   1st Qu.: 6.170  
 Median : 7.11   Median : 7.040  
 Mean   : 7.50   Mean   : 7.501  
 3rd Qu.: 7.98   3rd Qu.: 8.190  
 Max.   :12.74   Max.   :12.500  
```

Anscombe example (1973)

```
Analysis of Variance Table

Response: y1
          Df Sum Sq Mean Sq F value  Pr(>F)
x1         1 27.510 27.5100   17.99 0.00217 **
Residuals  9 13.763  1.5292
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Analysis of Variance Table

Response: y2
          Df Sum Sq Mean Sq F value   Pr(>F)
x2         1 27.500 27.5000  17.966 0.002179 **
Residuals  9 13.776  1.5307
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Analysis of Variance Table

Response: y3
          Df Sum Sq Mean Sq F value   Pr(>F)
x3         1 27.470 27.4700  17.972 0.002176 **
Residuals  9 13.756  1.5285
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Analysis of Variance Table

Response: y4
          Df Sum Sq Mean Sq F value   Pr(>F)
x4         1 27.490 27.4900  18.003 0.002165 **
Residuals  9 13.742  1.5269
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

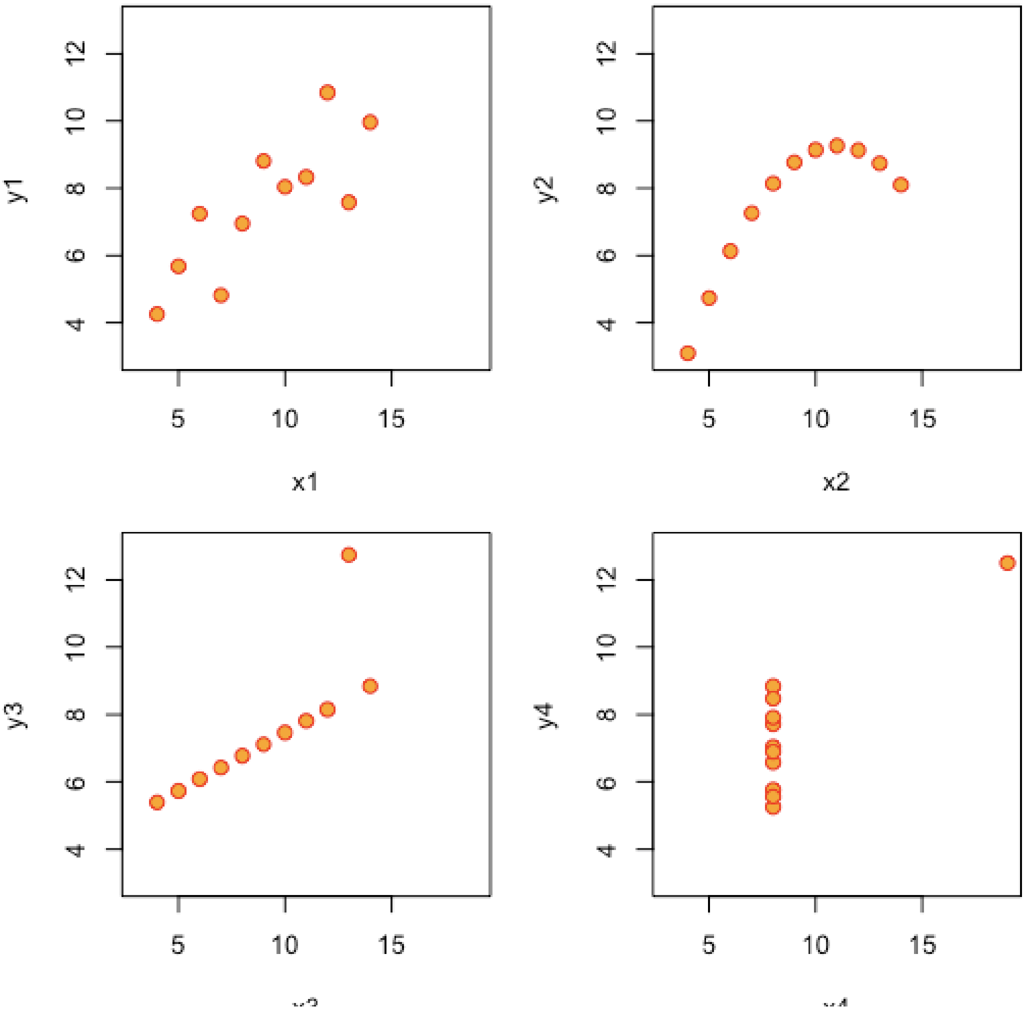
Anscombe example (1973)

	lm1	lm2	lm3	lm4
(Intercept)	3.0000909	3.000909	3.0024545	3.0017273
x1	0.5000909	0.500000	0.4997273	0.4999091

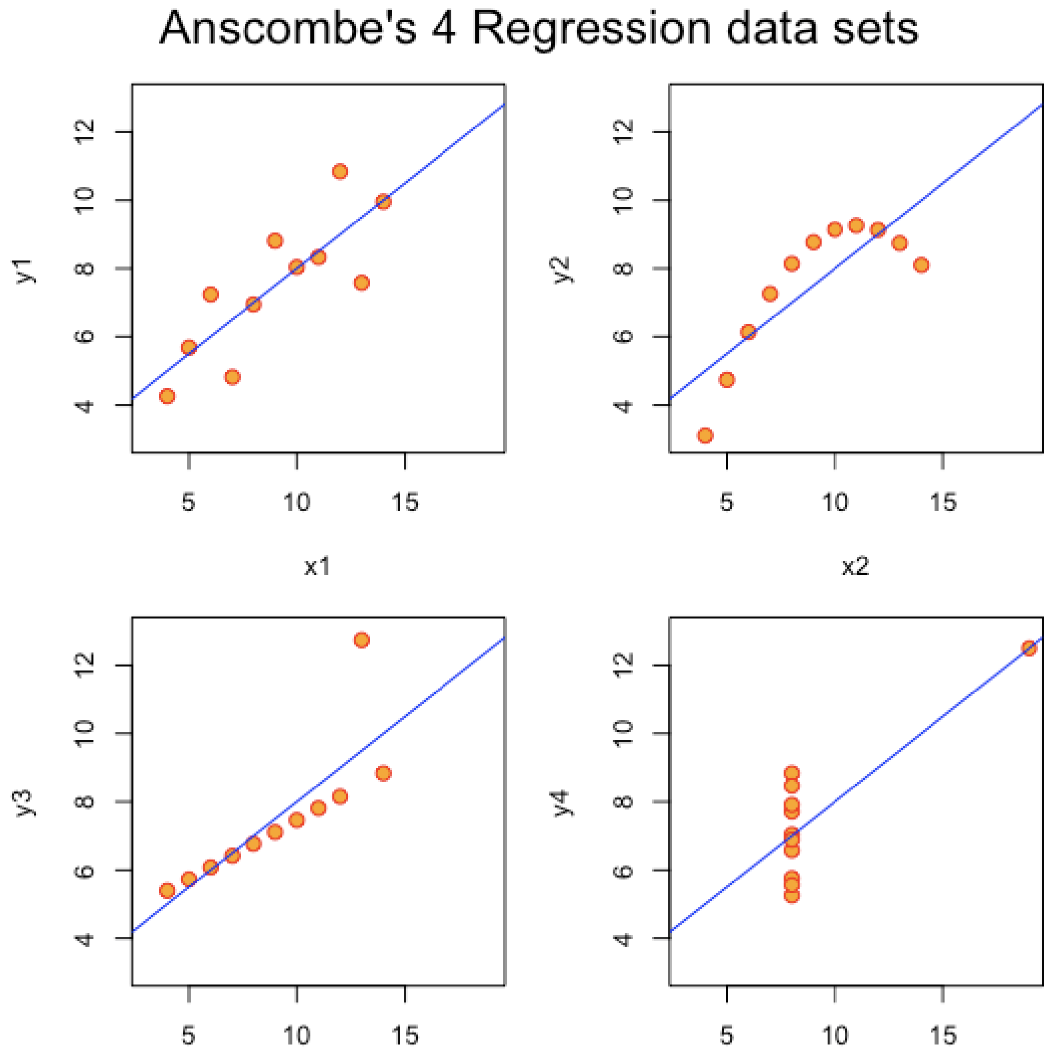
Anscombe example (1973)

```
$lm1
             Estimate Std. Error  t value    Pr(>|t|)
(Intercept) 3.0000909  1.1247468 2.667348 0.025734051
x1          0.5000909  0.1179055 4.241455 0.002169629

$lm2
            Estimate Std. Error  t value    Pr(>|t|)
(Intercept) 3.000909  1.1253024 2.666758 0.025758941
x2          0.500000  0.1179637 4.238590 0.002178816

$lm3
             Estimate Std. Error  t value    Pr(>|t|)
(Intercept) 3.0024545  1.1244812 2.670080 0.025619109
x3          0.4997273  0.1178777 4.239372 0.002176305

$lm4
             Estimate Std. Error  t value    Pr(>|t|)
(Intercept) 3.0017273  1.1239211 2.670763 0.025590425
x4          0.4999091  0.1178189 4.243028 0.002164602
```

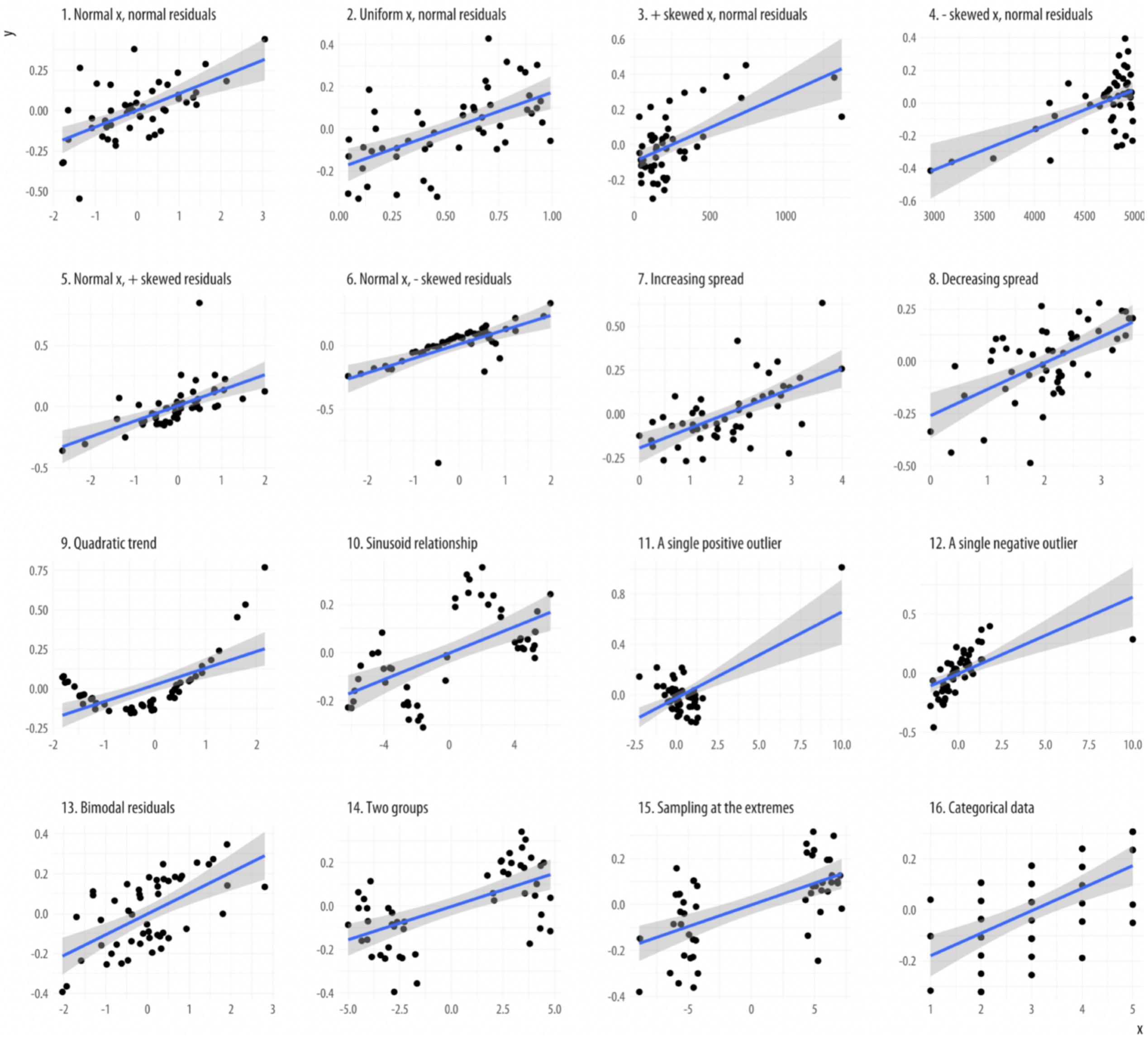
Tufte: Same relationship? (2001)



Tufte: Same relationship? (2001)



Jan Vanhove example (2016)



Elements of a Chart

Elements of a Chart

1. Dimensionality
    1. How many dimensions are there?
2. Relationships
    1. Strength
    2. Fit
    3. Error bands
    4. Panels

Figure 1.8: A crisis of faith in democracy? (New York Times.)

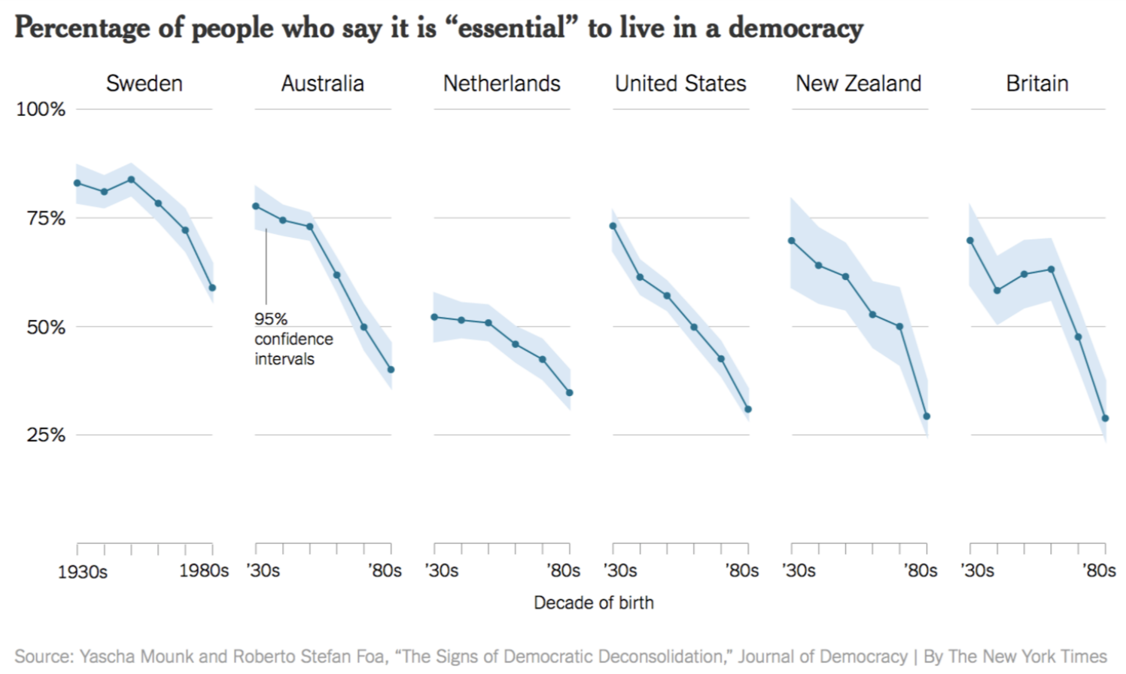


Figure 1.9: Perhaps the crisis has been overblown. (Erik Voeten.)

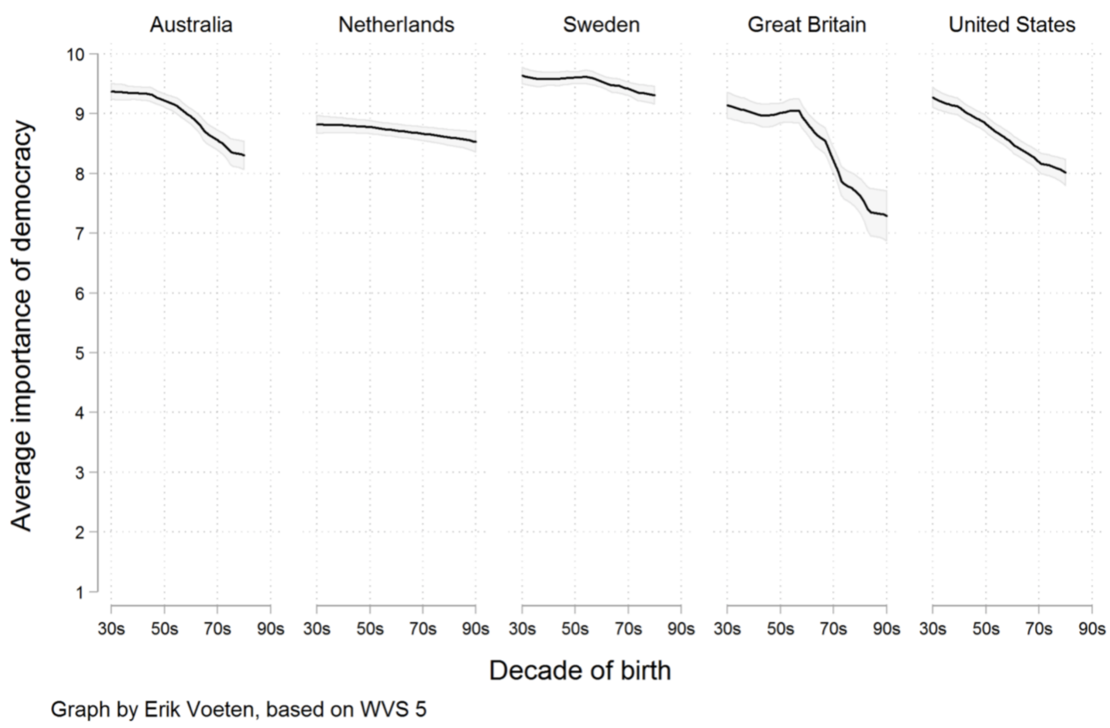
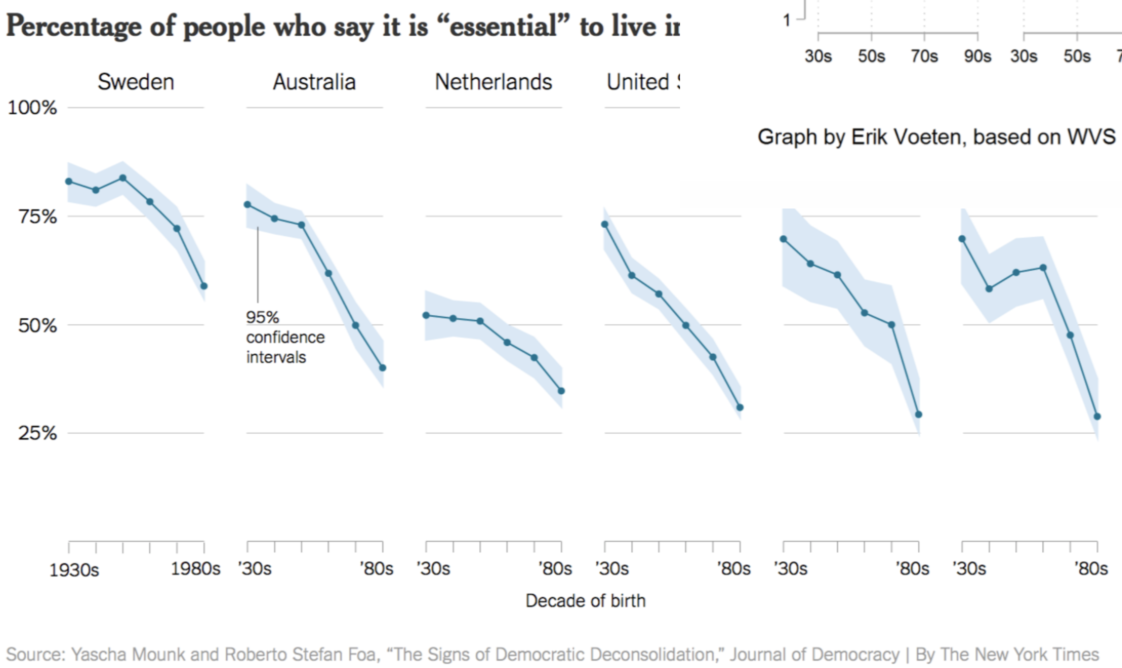


Figure 1.8: A crisis of faith in democracy? (New York



Scale matters!

Figure 1.9: Perhaps the crisis has been overblown. (Erik Voeten.)

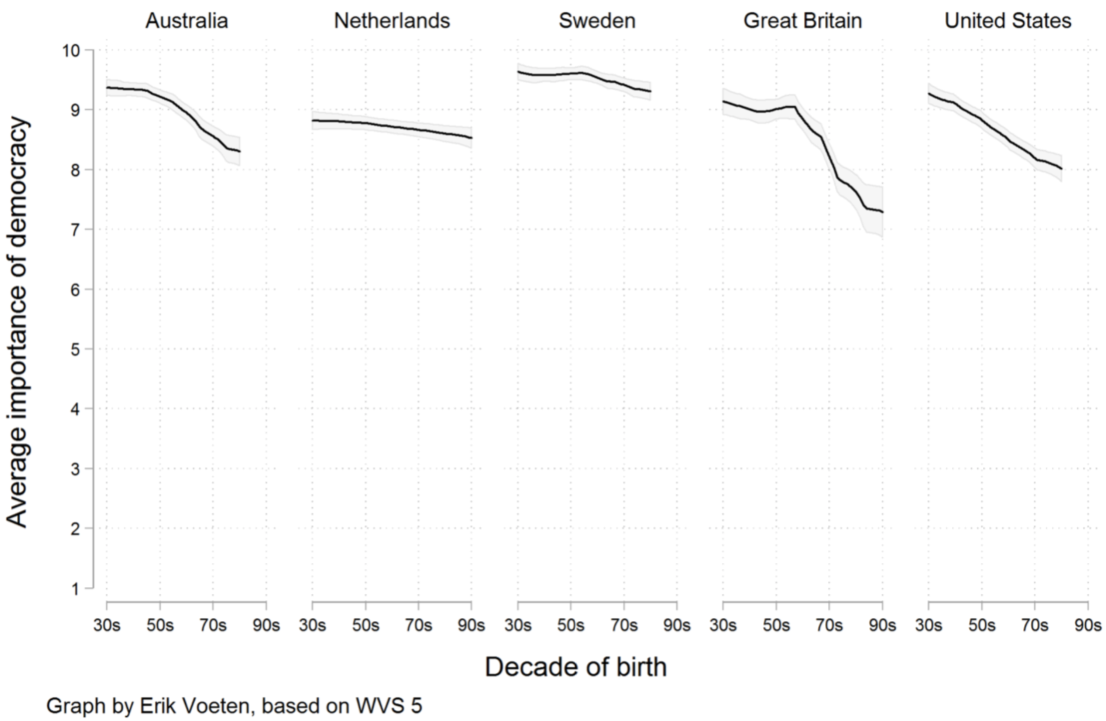
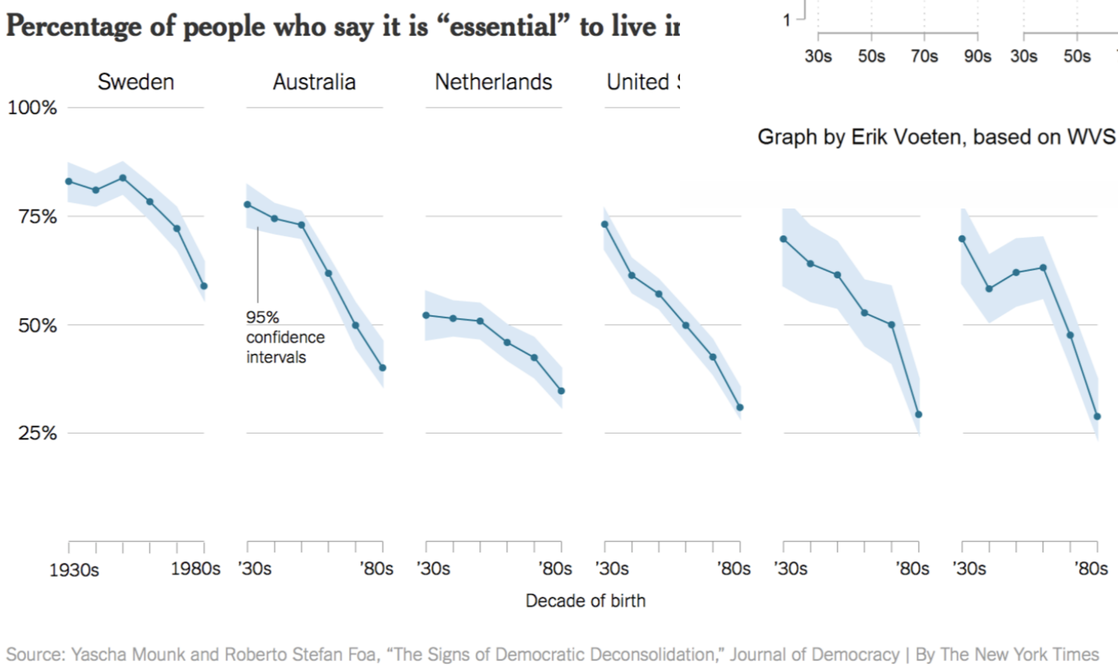


Figure 1.8: A crisis of faith in democracy? (New York



Take-away

Take-away

In a data chart, there is always something we can do better.

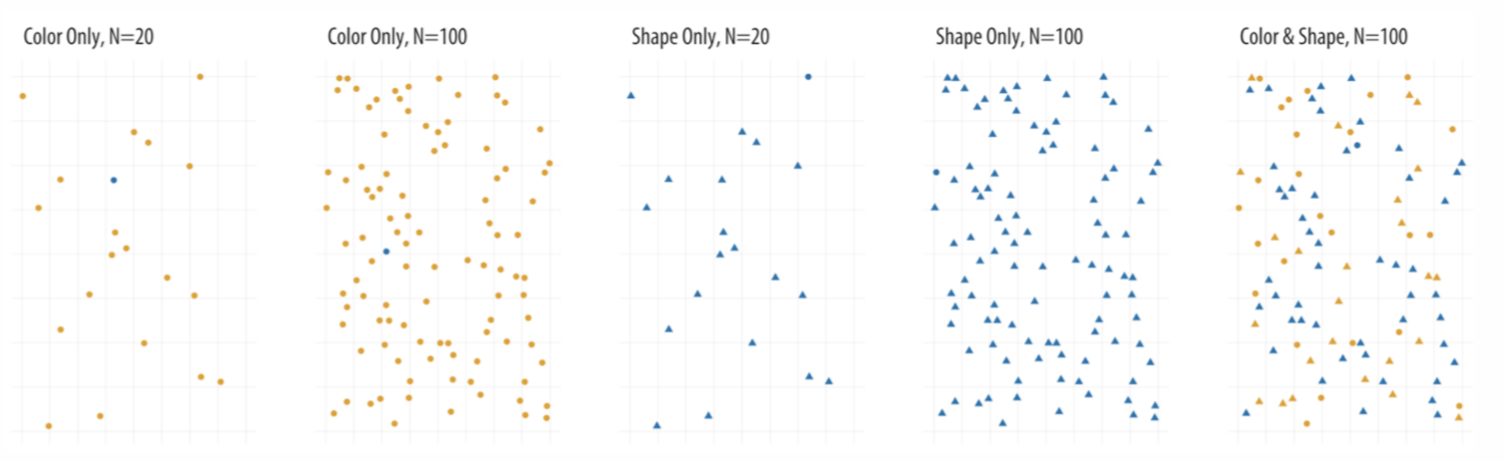
Pre-attentive pop-out

Pre-attentive pop-out

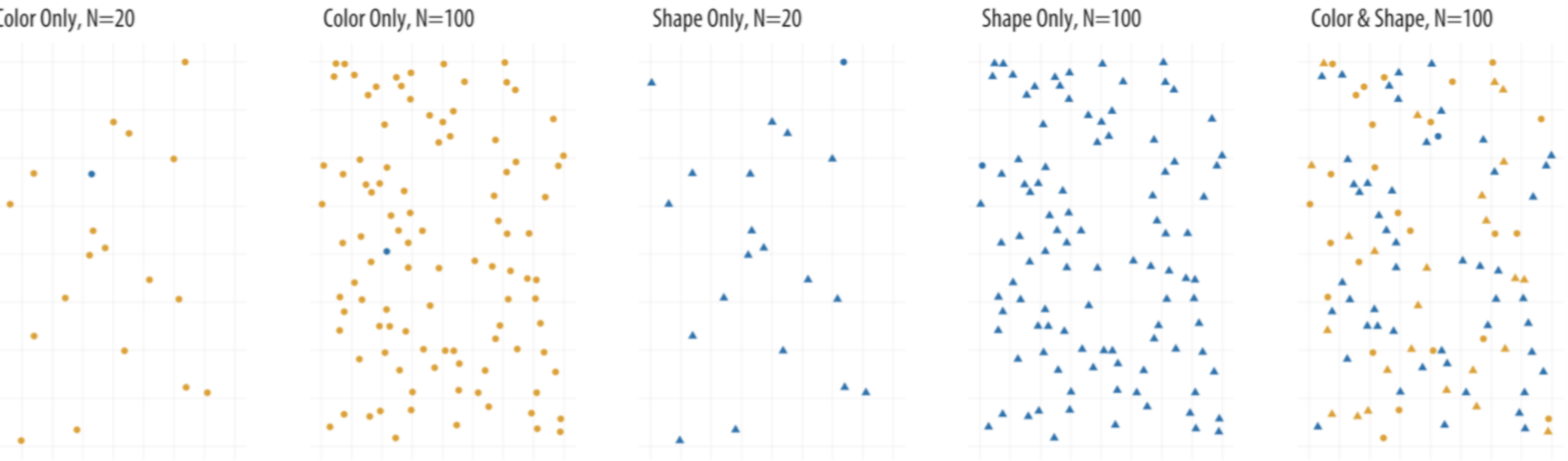
Some objects in our visual field are easier to see than others (e.g. color, shapes).

Pre-attentive pop-out

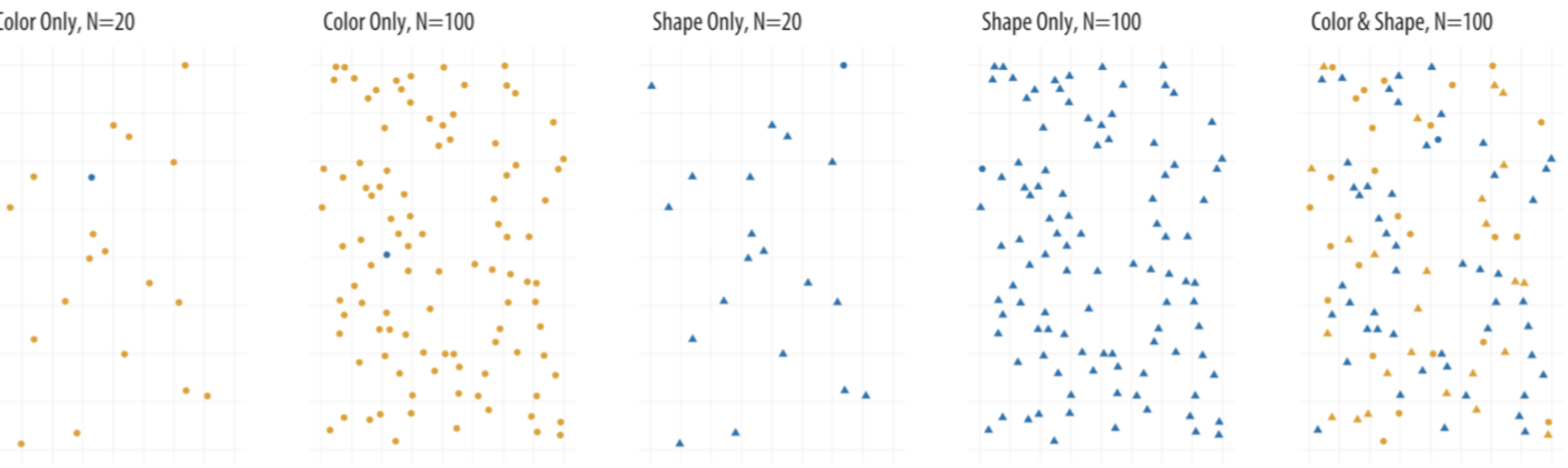
Some objects in our visual field are easier to see than others (e.g. color, shapes).



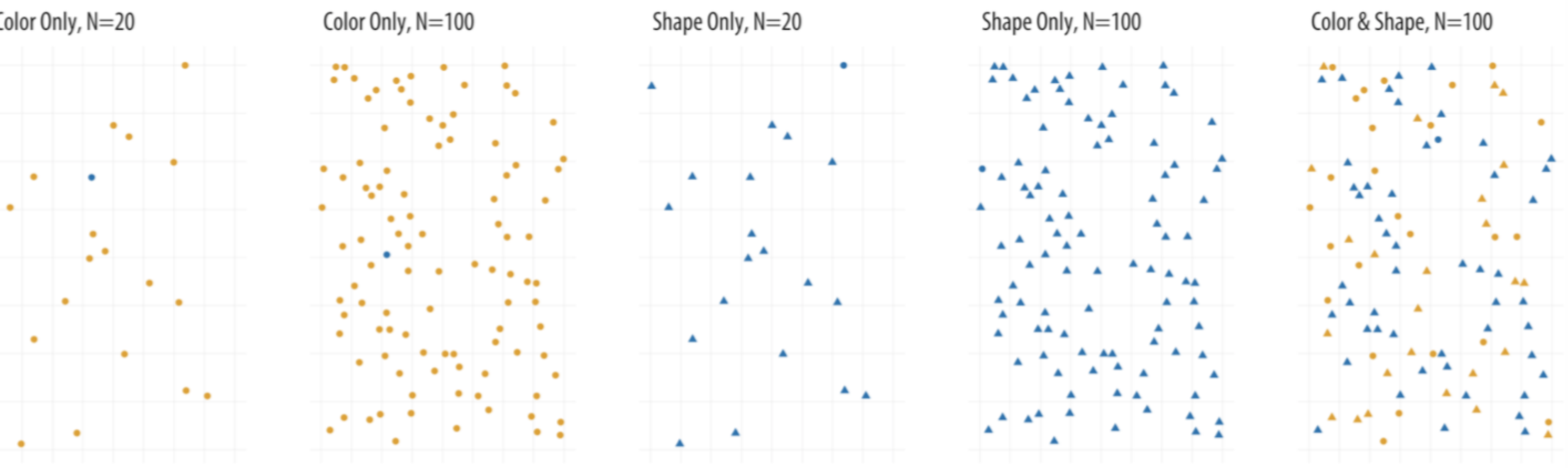
Pre-attentive pop-out



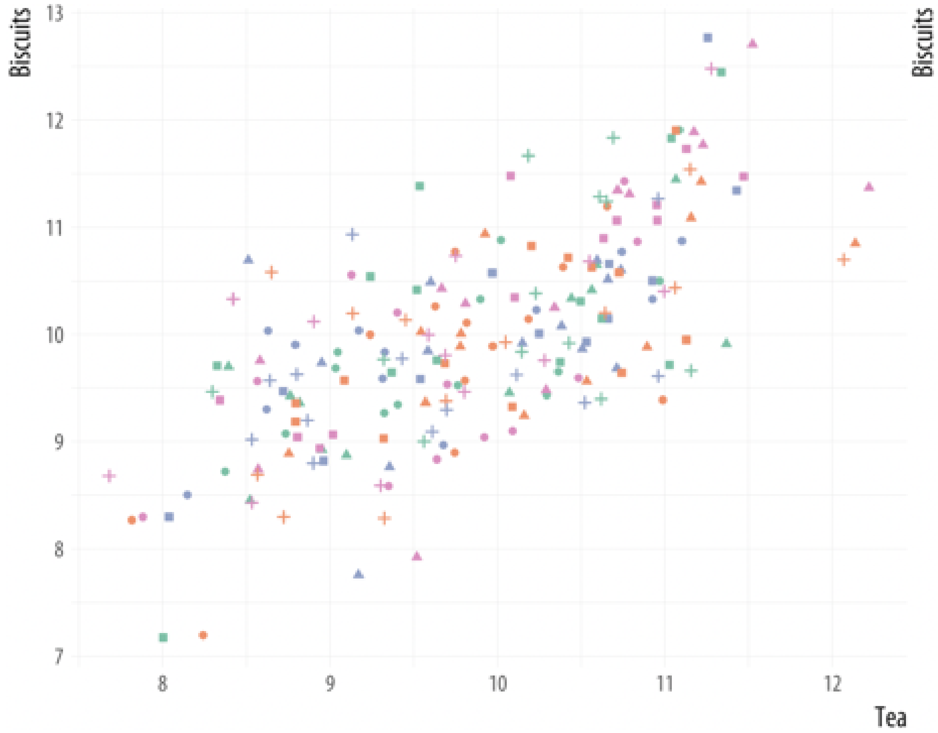
Pre-attentive pop-out

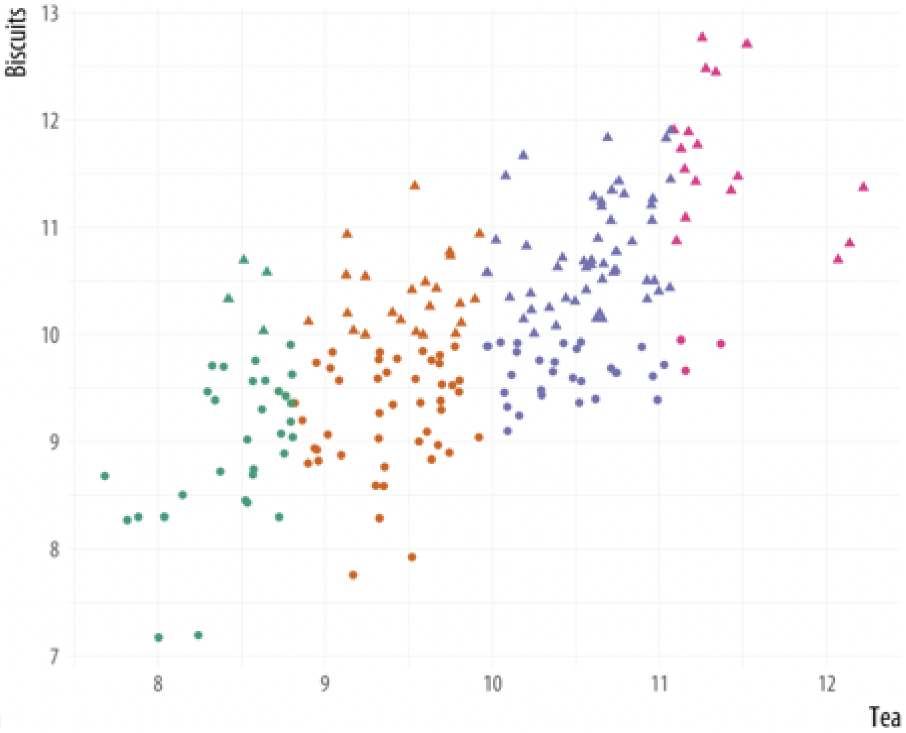


Pre-attentive pop-out

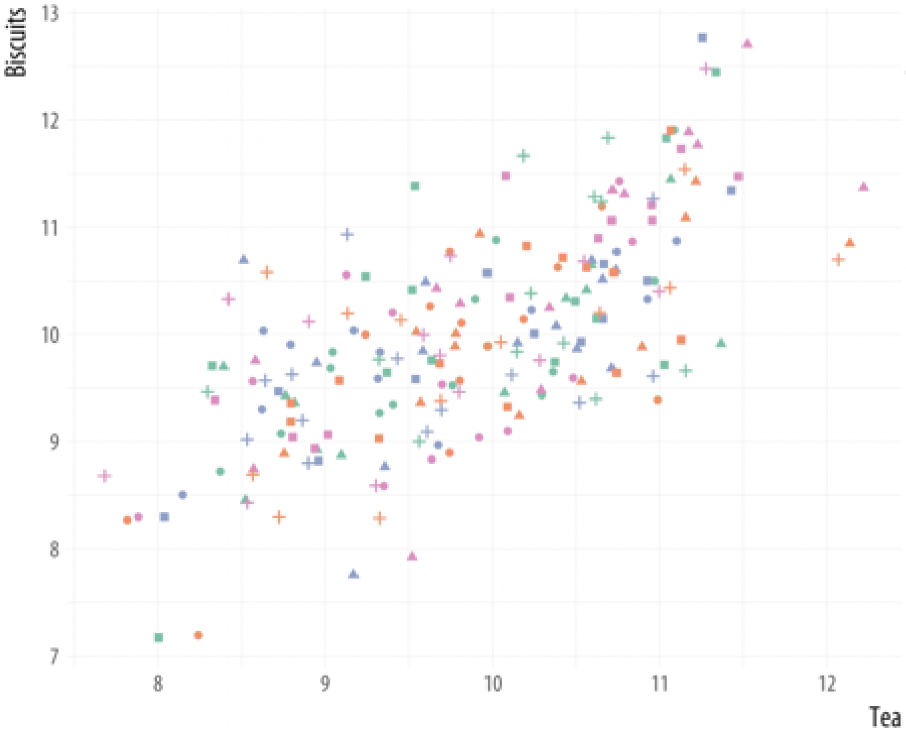


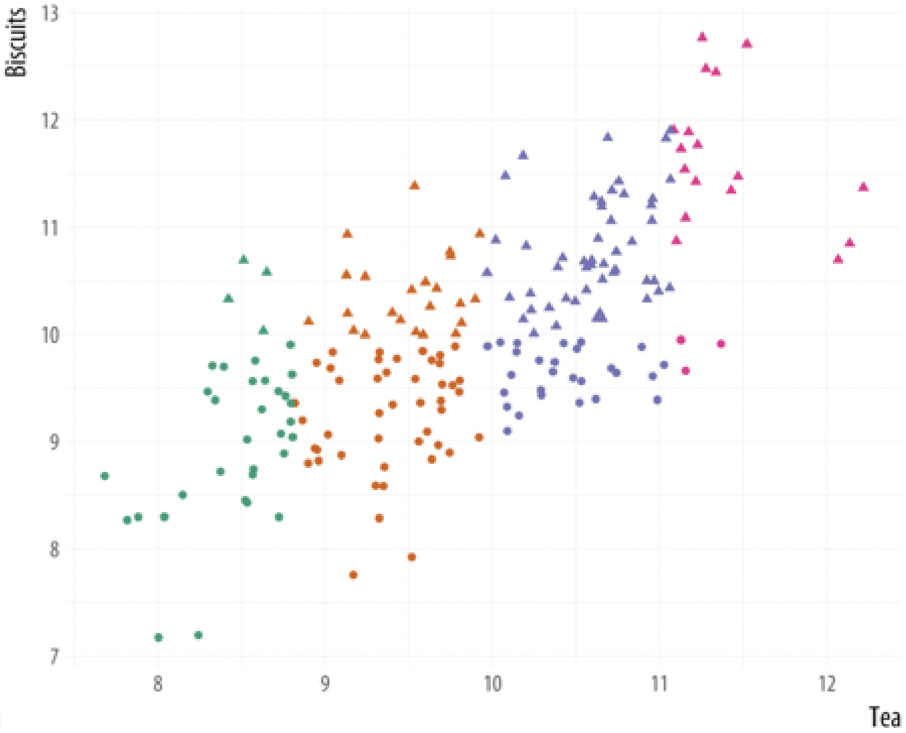
Pre-attentive pop-out





Pre-attentive pop-out





Even harder. Why?

Gestalt Rules

Gestalt Rules

The strong inferences we make about relationships between visual elements from relatively sparse visual information to identify groupings, classifications, or entities.

Gestalt Rules

Gestalt Rules

1. Proximity

Gestalt Rules

1. Proximity
2. Similarity

Gestalt Rules

1. Proximity
2. Similarity
3. Connection

Gestalt Rules

1. Proximity
2. Similarity
3. Connection
4. Continuity

Gestalt Rules

1. Proximity
2. Similarity
3. Connection
4. Continuity
5. Closure

Gestalt Rules

1. Proximity
2. Similarity
3. Connection
4. Continuity
5. Closure
6. Figure and ground

Gestalt Rules

1. Proximity
2. Similarity
3. Connection
4. Continuity
5. Closure
6. Figure and ground
7. Common fate

Gestalt Rules

Gestalt Rules

1. Proximity

   Things that are spatially near to one another seem to be related.

Gestalt Rules

1. Proximity

   Things that are spatially near to one another seem to be related.

2. Similarity

   Things that look alike seem to be related.

Gestalt Rules

Gestalt Rules

3. Connection

   Things that are visually tied to one another seem to be related.

Gestalt Rules

3. Connection

   Things that are visually tied to one another seem to be related.

4. Continuity

   Partially hidden objects are completed into familiar shapes.

Gestalt Rules

Gestalt Rules

5. Closure:

   Incomplete shapes are perceived as complete.

Gestalt Rules

5. Closure:

   Incomplete shapes are perceived as complete.

6. Figure and Ground

   Visual elements are taken to be either in the foreground or the background.

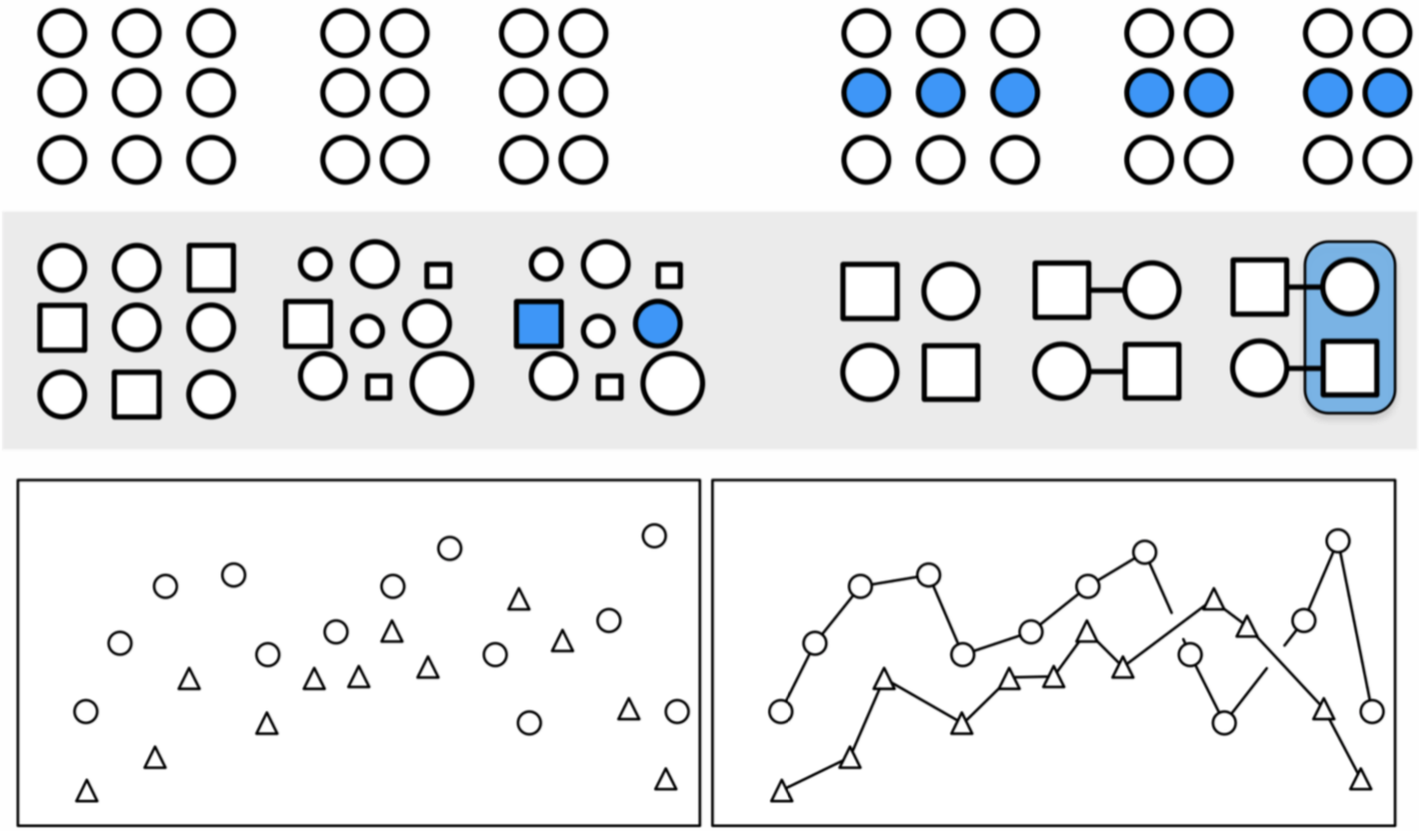
Gestalt Rules

Gestalt Rules

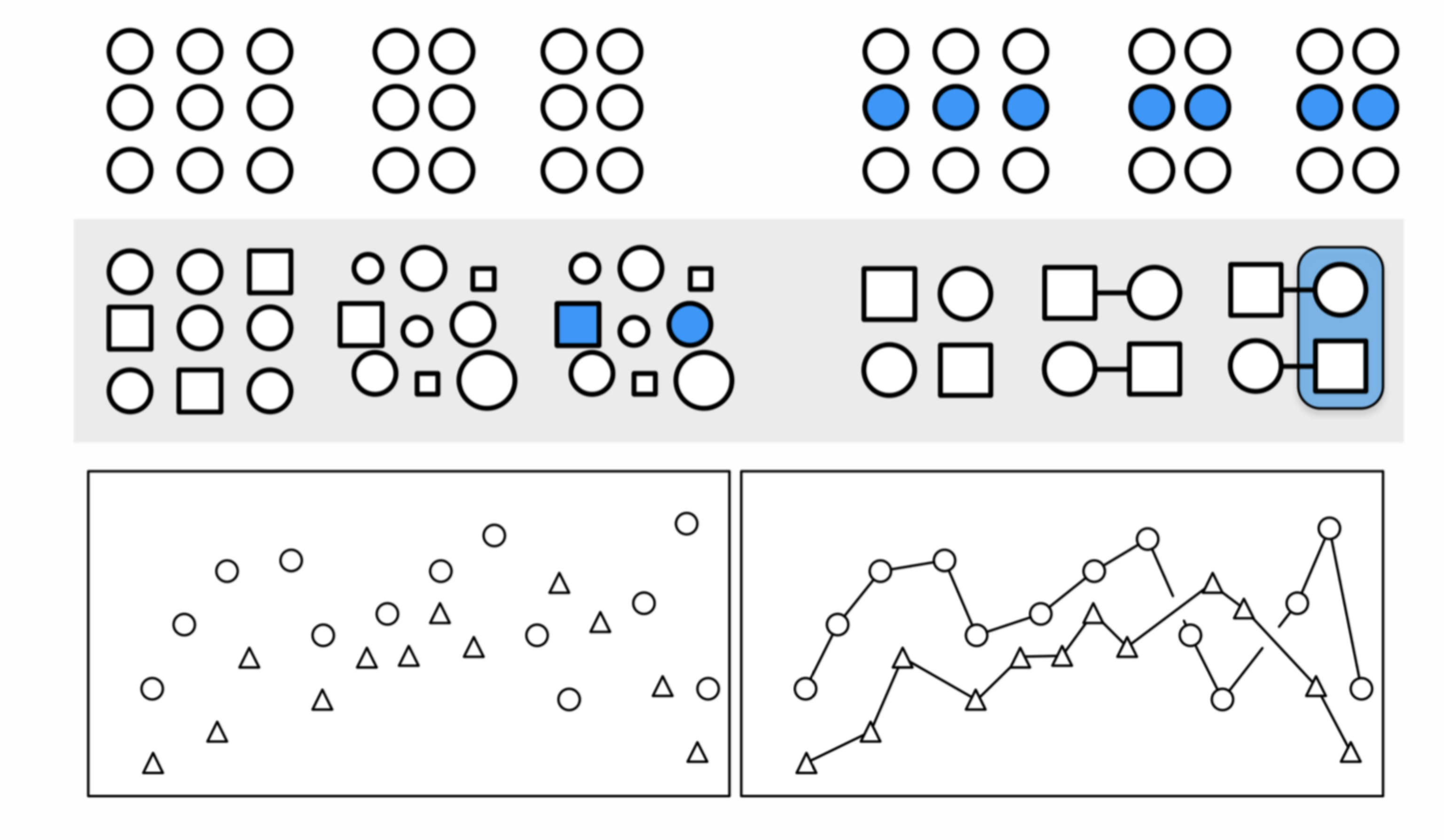
7. Common Fate:

   Elements sharing a direction of movement are perceived as a unit.

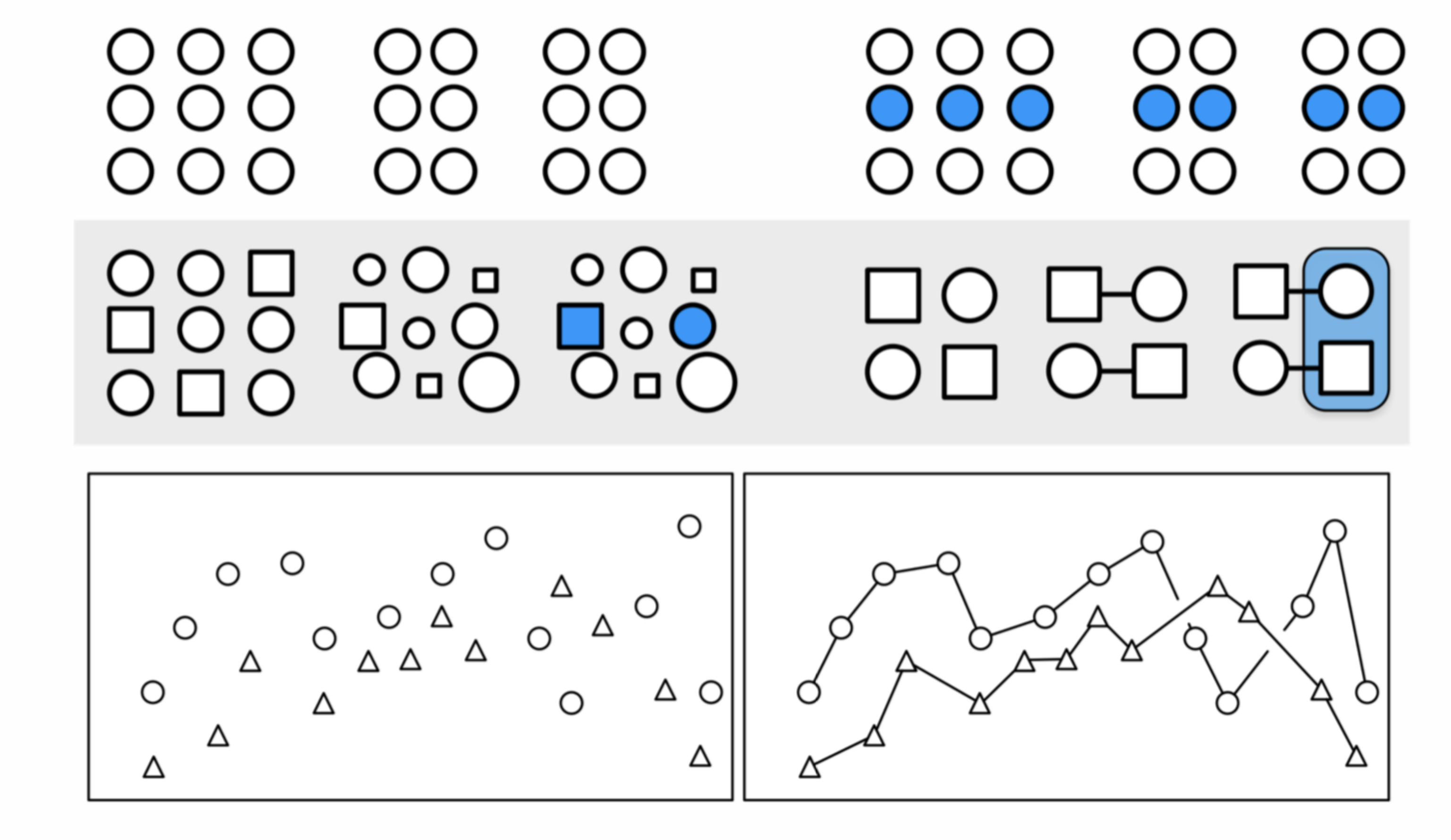
Gestalt Rules



Gestalt Rules



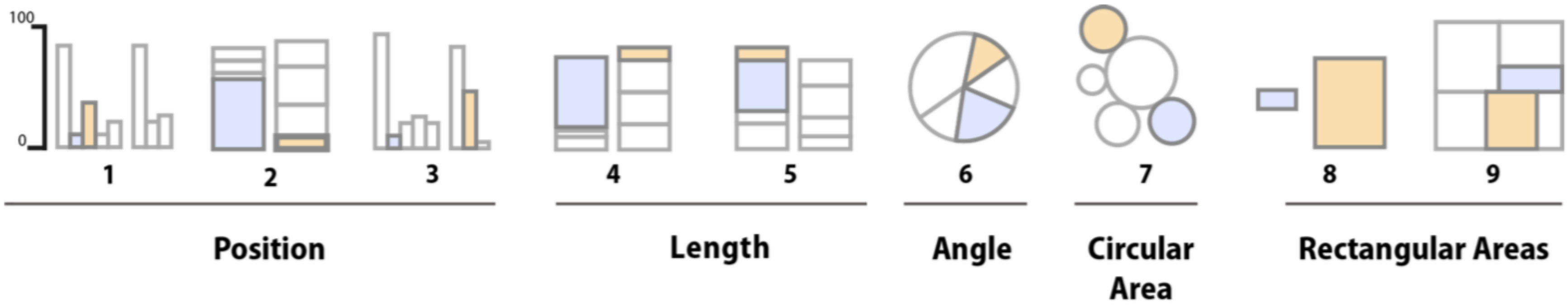
Gestalt Rules



How many groups/dimensions in each?

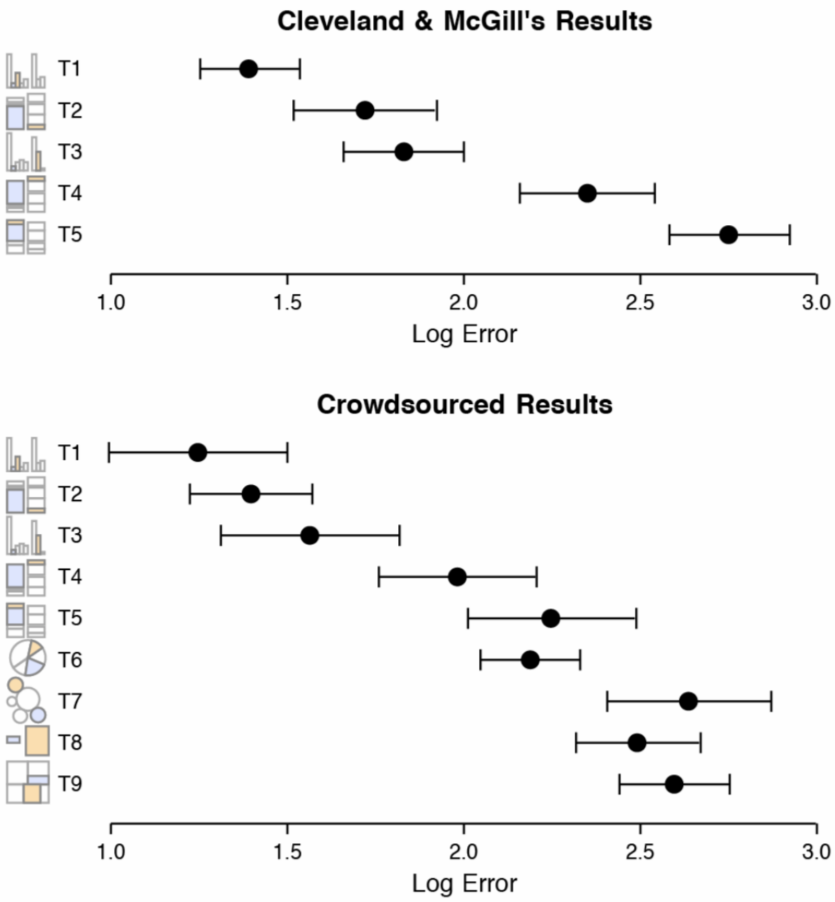
Cleveland and McGill Experiments

Cleveland and McGill Experiments

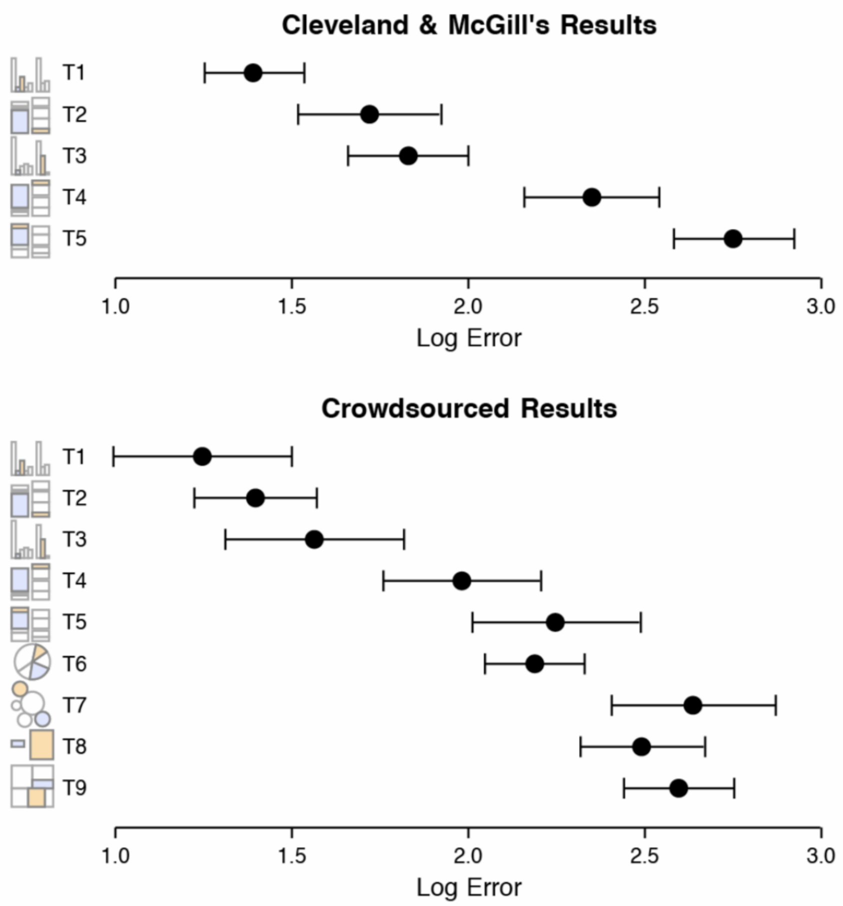


Cleveland and McGill Experiments

Cleveland and McGill Experiments



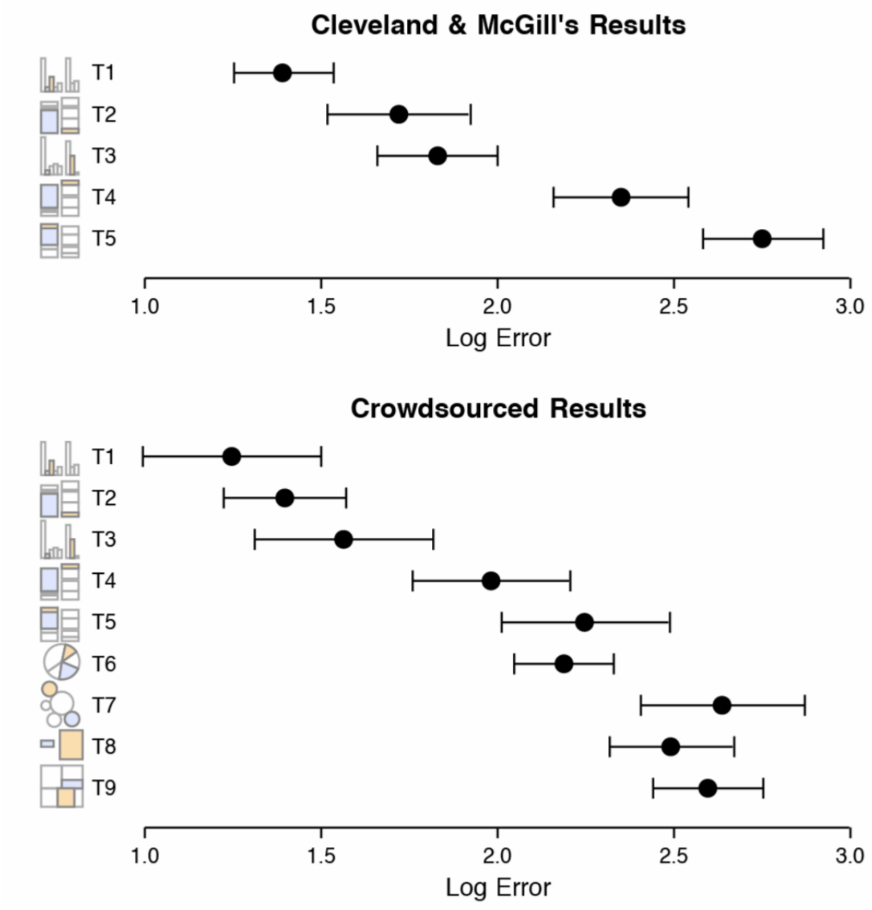
Cleveland and McGill Experiments



Comparison:

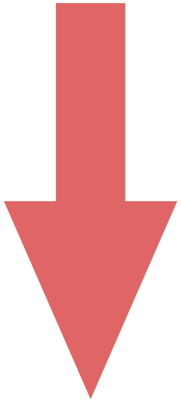
- relative position
- on a common scale
- by length
- by angles
- by areas
- by volume

Cleveland and McGill Experiments



Comparison:

- relative position
- on a common scale
- by length
- by angles
- by areas
- by volume



Miller's Law

Miller's Law

"The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information"

Miller's Law

"The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information"

- George Miller 1956

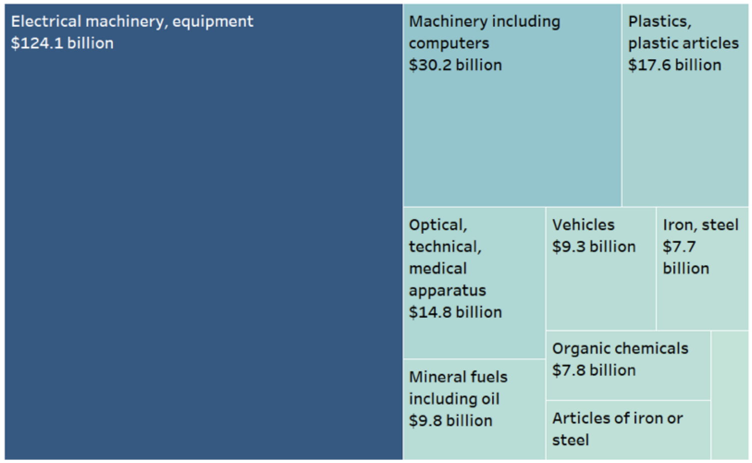
R Graphics

At the heart of the R graphics facilities is the package *grDevices*, which will be referred to as the graphics engine. This provides fundamental infrastucture for graphics in R, such as selecting colors and fonts and selecting a graphics output format.

Showcase:

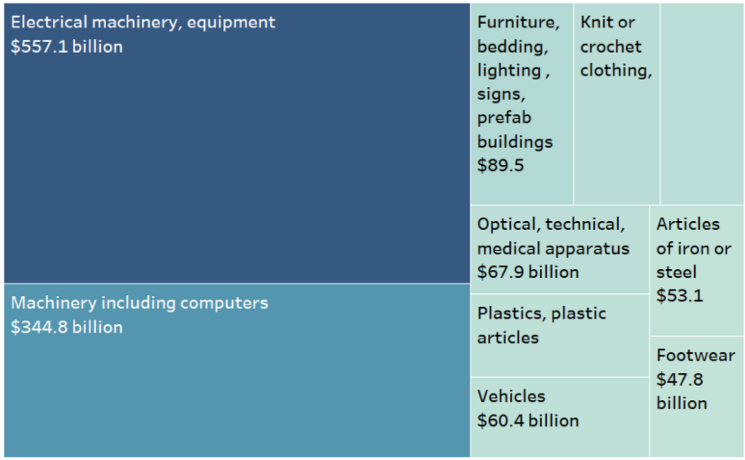
Ho, Clark and Tan 2018

Taiwan's Top 10 Exports by Product Groups



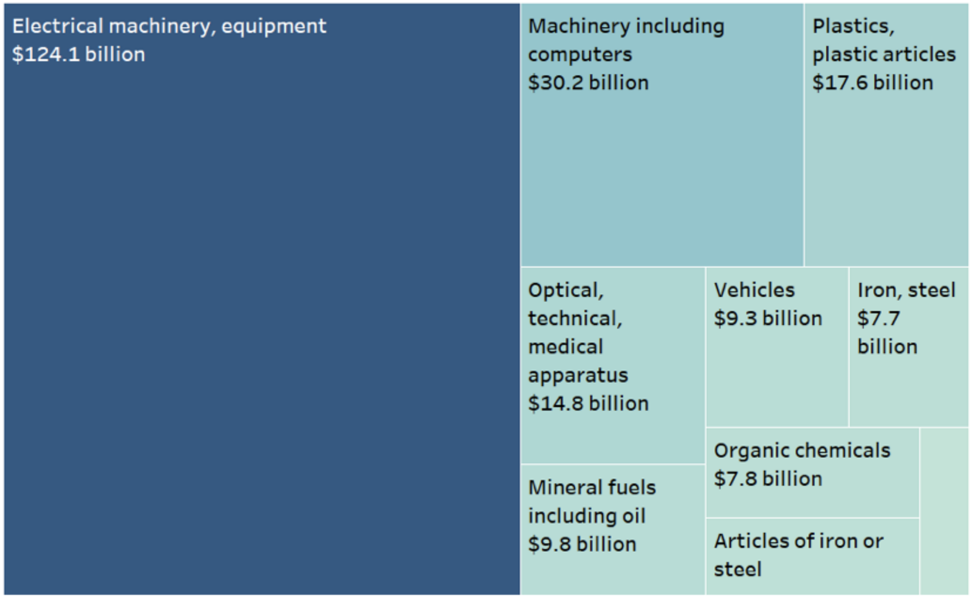
F1 and F2. Color shows sum of F3. Size shows sum of F3. The marks are labeled by F1 and F2.

China's Top 10 Exports by Product Groups



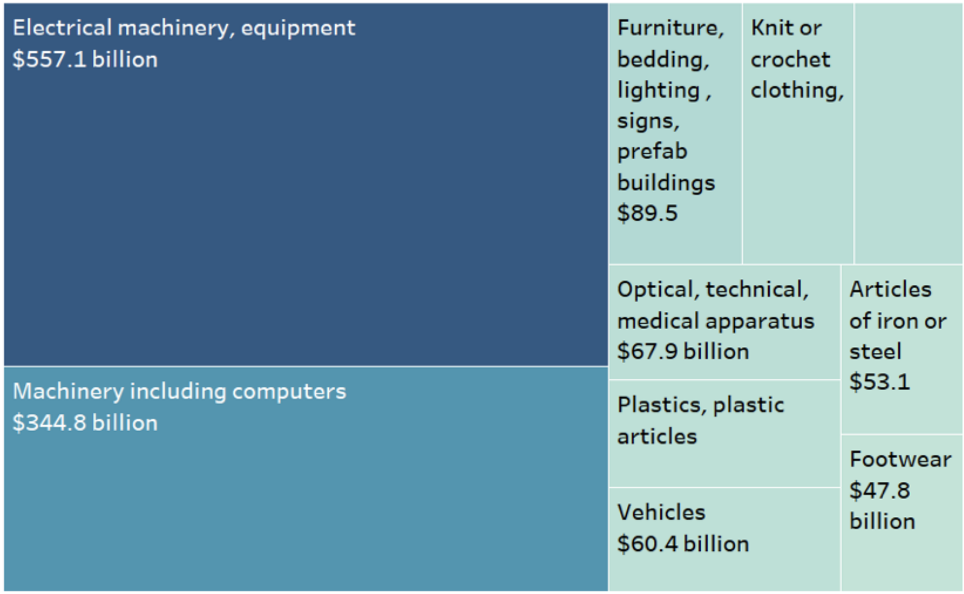
F1 and F3. Color shows sum of F2. Size shows sum of F2. The marks are labeled by F1 and F3.

Taiwan's Top 10 Exports by Product Groups



F1 and F2. Color shows sum of F3. Size shows sum of F3. The marks are labeled by F1 and F2.

China's Top 10 Exports by Product Groups



F1 and F3. Color shows sum of F2. Size shows sum of F2. The marks are labeled by F1 and F3.

Showcase:

A. Q. Philips 2018 (AJPS)

FIGURE 1 The ARDL-Bounds Procedure's Comprehensive Approach to Time-Series Analysis

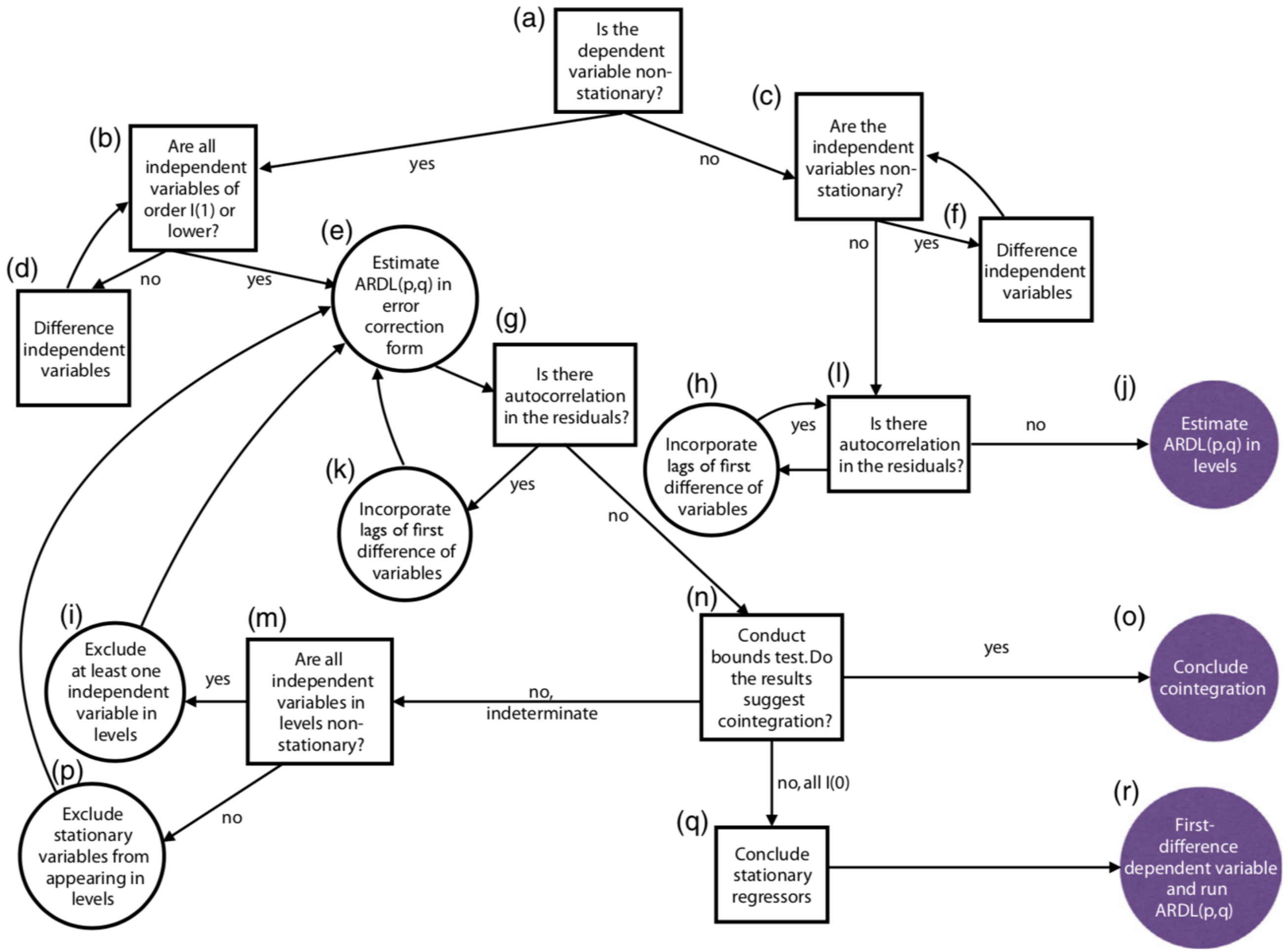
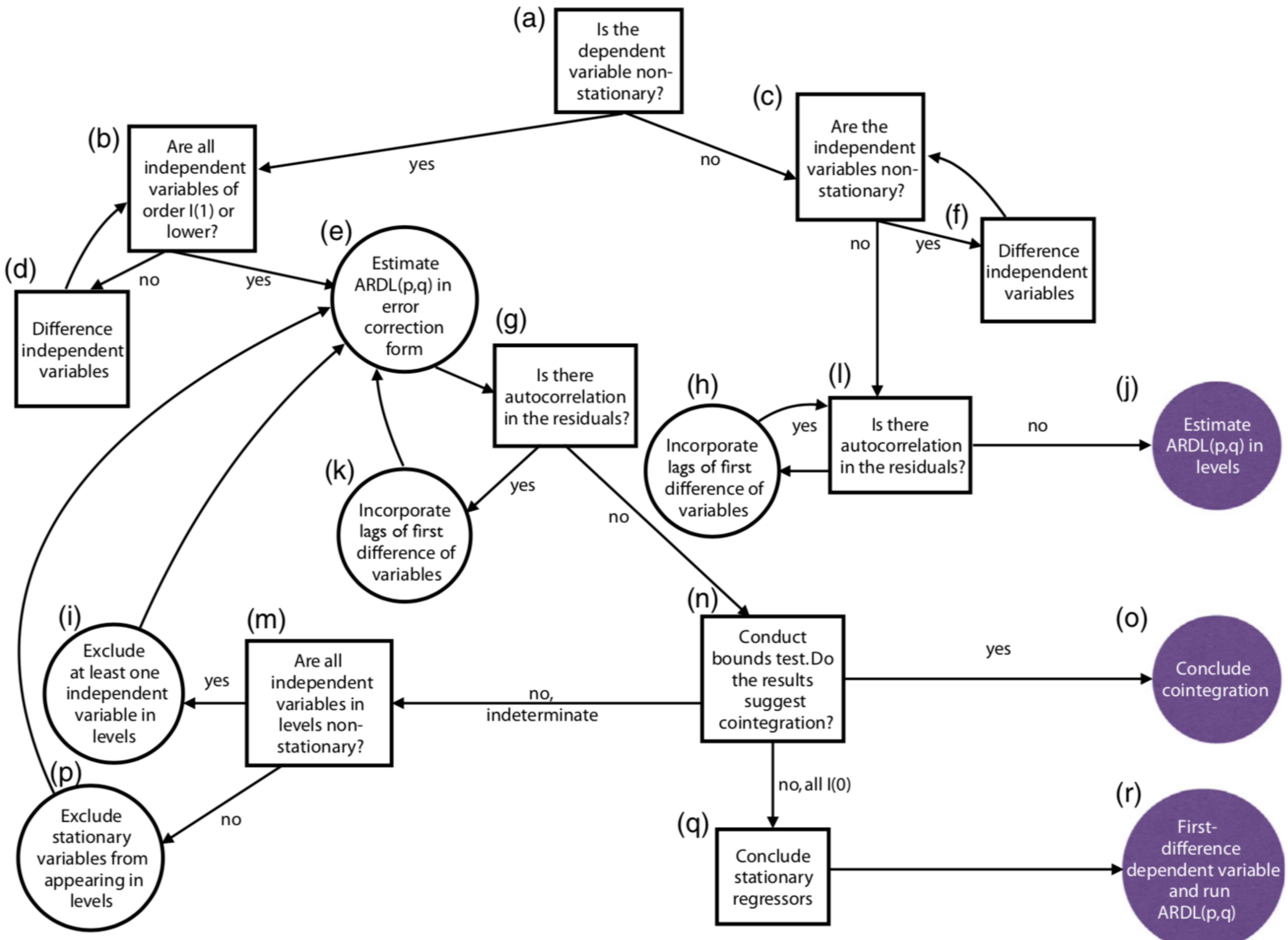


FIGURE 1 The ARDL-Bounds Procedure's Comprehensive Approach to Time-Series Analysis



R Graphics

Two packages build directly on top of the graphics engine:

1. graphics package
2. grid package

These represent two largely incompatible graphics systems and they divide the bulk of graphics functionality in R into two separate worlds.

Rule # 1

Rule # 1

Never use default.

If you're not doing something different, you're not doing anything at all.

If you’re not doing something different, you’re not doing anything at all.

It comes with the package, to reconcile [yourself] to life’s inevitable trade-offs and heartaches.

If you’re not doing something different, you’re not doing anything at all.

It comes with the package, to reconcile [yourself] to life’s inevitable trade-offs and heartaches.

- Edward Tufte

Brainstormer

Brainstormer

Journalistic Data Visualization

vs.

Scientific Data Visualization

What do you see?

What do you see?

What do you see?

Cognitive Scientist:
The Visualization Process

Cognitive Scientist: The Visualization Process

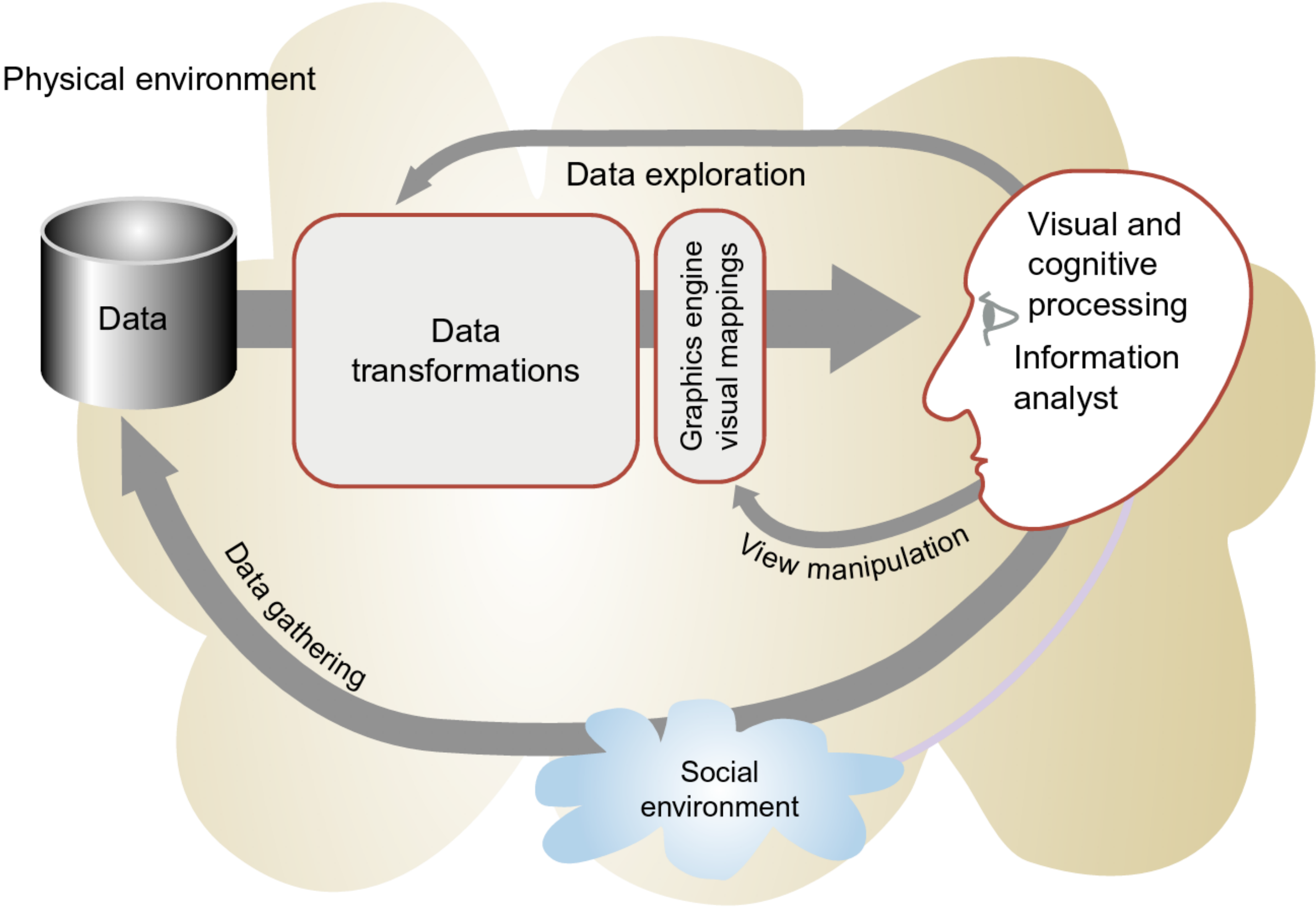


Figure 1.2 The visualization process.

Source: Ware, Colin 2012.

Cognitive Scientist: The Visualization Process

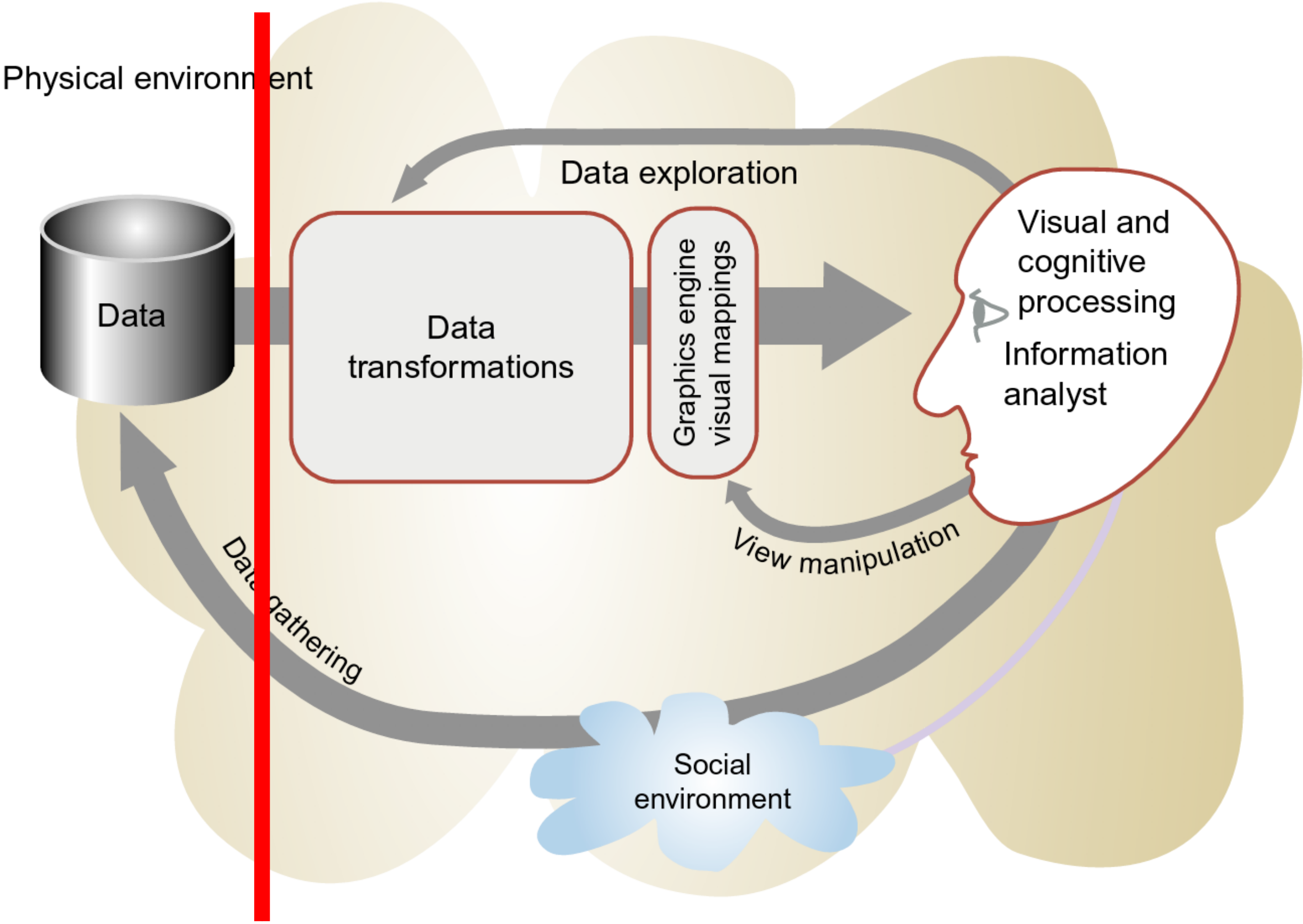


Figure 1.2 The visualization process.

Source: Ware, Colin 2012.

Cognitive Scientist: The Visualization Process

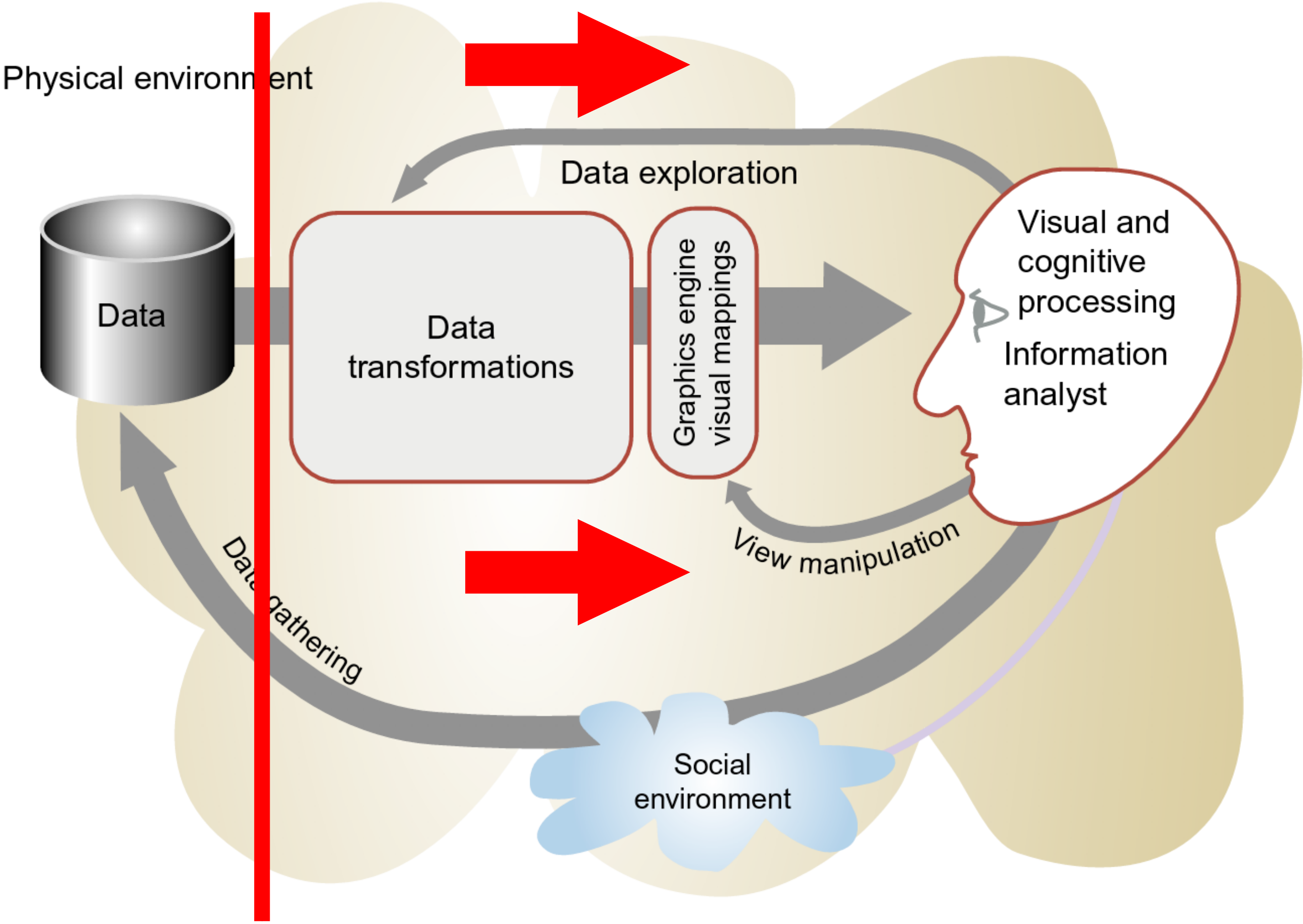


Figure 1.2 The visualization process.

Source: Ware, Colin 2012.

Cognitive Scientist: The Visualization Process

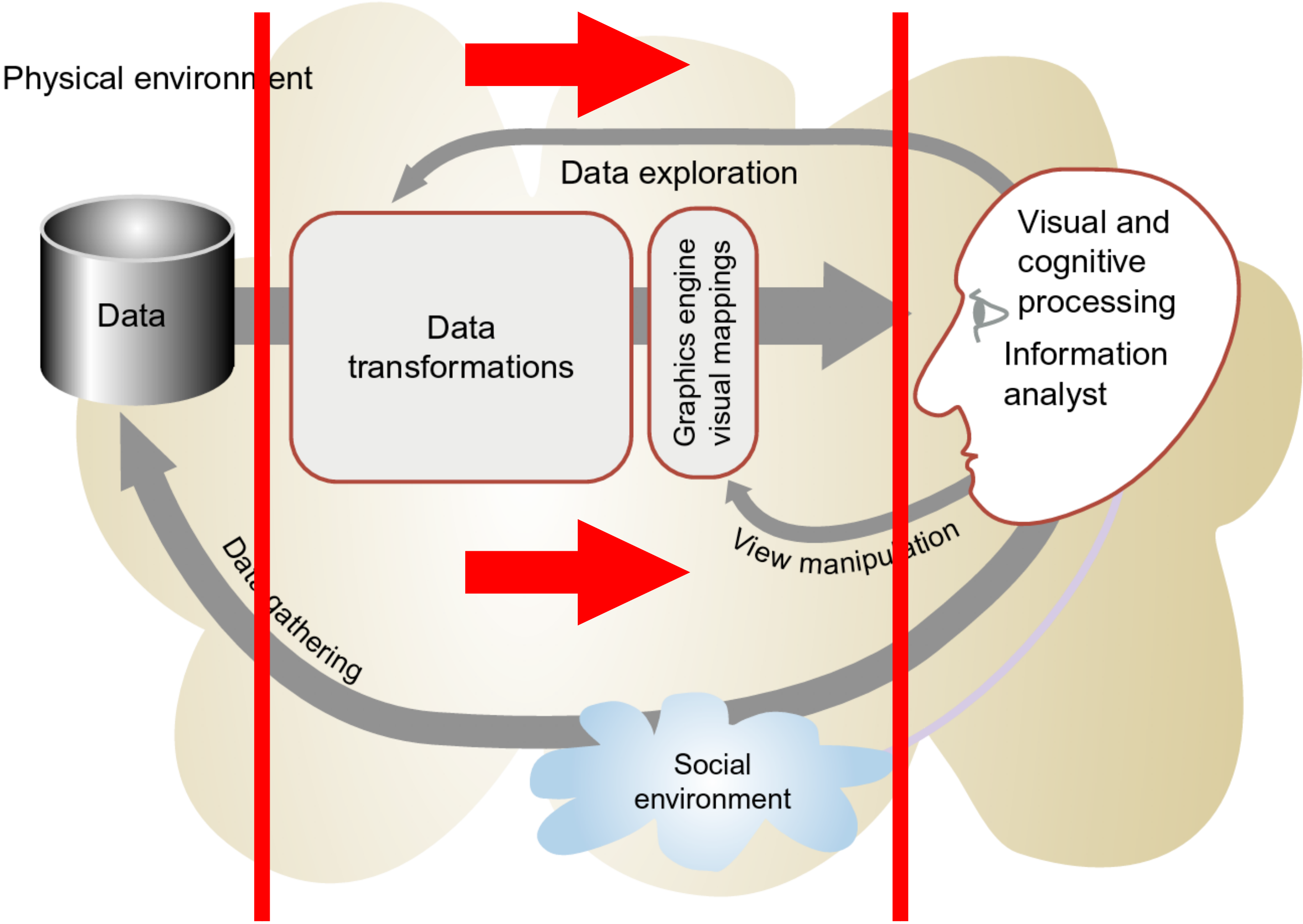


Figure 1.2 The visualization process.

Source: Ware, Colin 2012.

Visualization Stages

Visualization Stages

1. The collection and storage of data.

Visualization Stages

1. The collection and storage of data.
2. A preprocessing stage designed to transform the data into something that is easier to manipulate.

Visualization Stages

1. The collection and storage of data.
2. A preprocessing stage designed to transform the data into something that is easier to manipulate.
3. Mapping from the selected data to a visual representation, which is accomplished through computer algorithms that produce an image on the screen.

Visualization Stages

1. The collection and storage of data.
2. A preprocessing stage designed to transform the data into something that is easier to manipulate.
3. Mapping from the selected data to a visual representation, which is accomplished through computer algorithms that produce an image on the screen.
4. The human perceptual and cognitive system (the perceiver).

Data Visualization is:

Data Visualization is:

....the revelation of the complex.

Data Visualization is:

....the revelation of the complex.

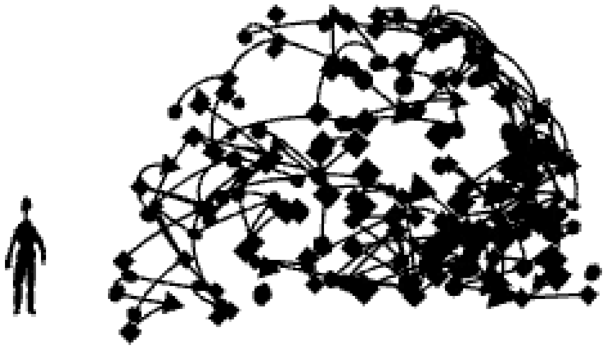
- Tufte 2001

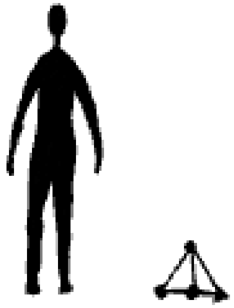
What is Complexity?

What is Complexity?

Source: Ammer, Ralph. 2017. "Make me think!" *Medium* https://blog.prototypr.io/make-me-think-90b46aa50513

What is Simplicity?





Source: Ammer, Ralph. 2017. "Make me think!" *Medium* https://blog.prototypr.io/make-me-think-90b46aa50513

Should the technology grow?

Should the technology grow?

Or should the user?

Should the technology grow?

Or should the user?

- Ammer 2018

Which two animals have very similar visual structures with humans?

Which two animals have very similar visual structures with humans?

Monkeys and cats.

Anatomically, we all have the same visual system, i.e. it is likely that we all see in the same way.

Anatomically, we all have the same visual system, i.e. it is likely that we all see in the same way.

Visual designs should apply for all of us,

Anatomically, we all have the same visual system, i.e. it is likely that we all see in the same way.

Visual designs should apply for all of us, except color-blindness

Cognitive Scientist:
Visual Information Processing

Cognitive Scientist: Visual Information Processing

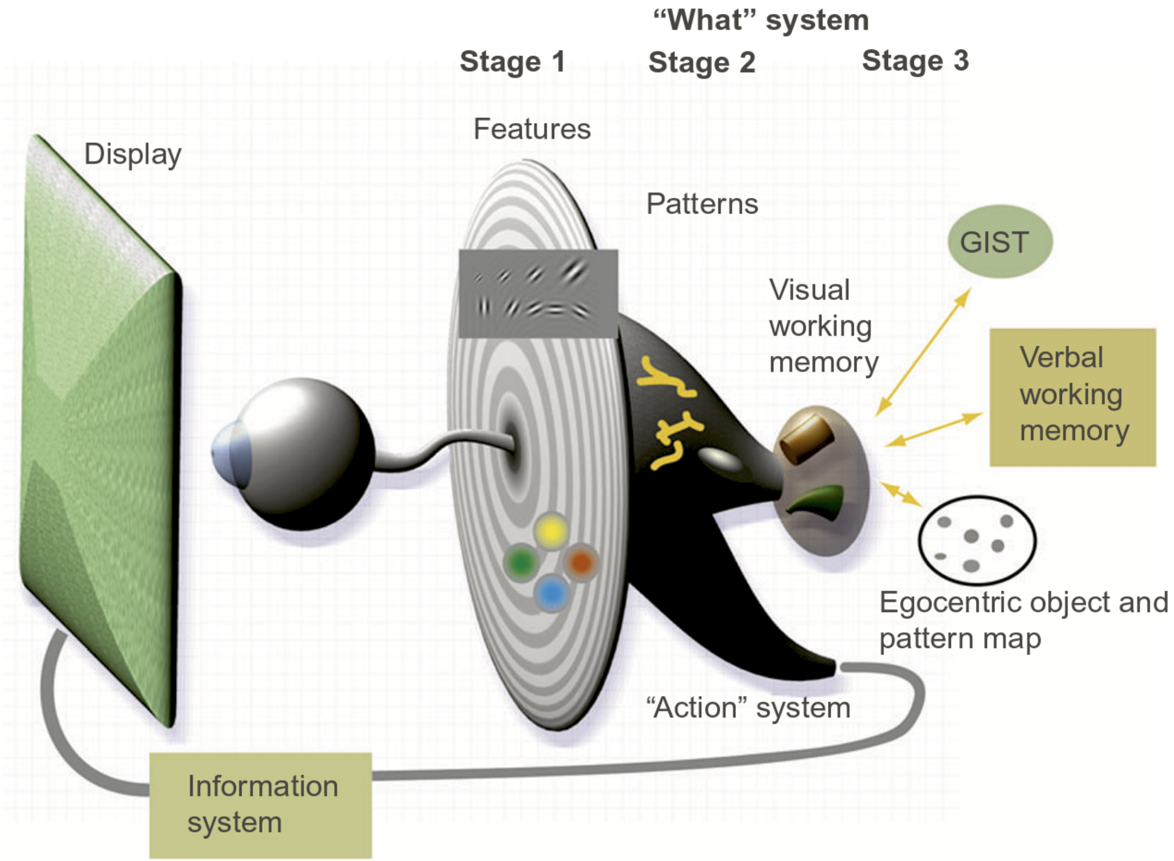


Figure 1.11 A three-stage model of visual information processing.

Source: Ware, Colin 2012.

Cognitive Scientist: Visual Information Processing

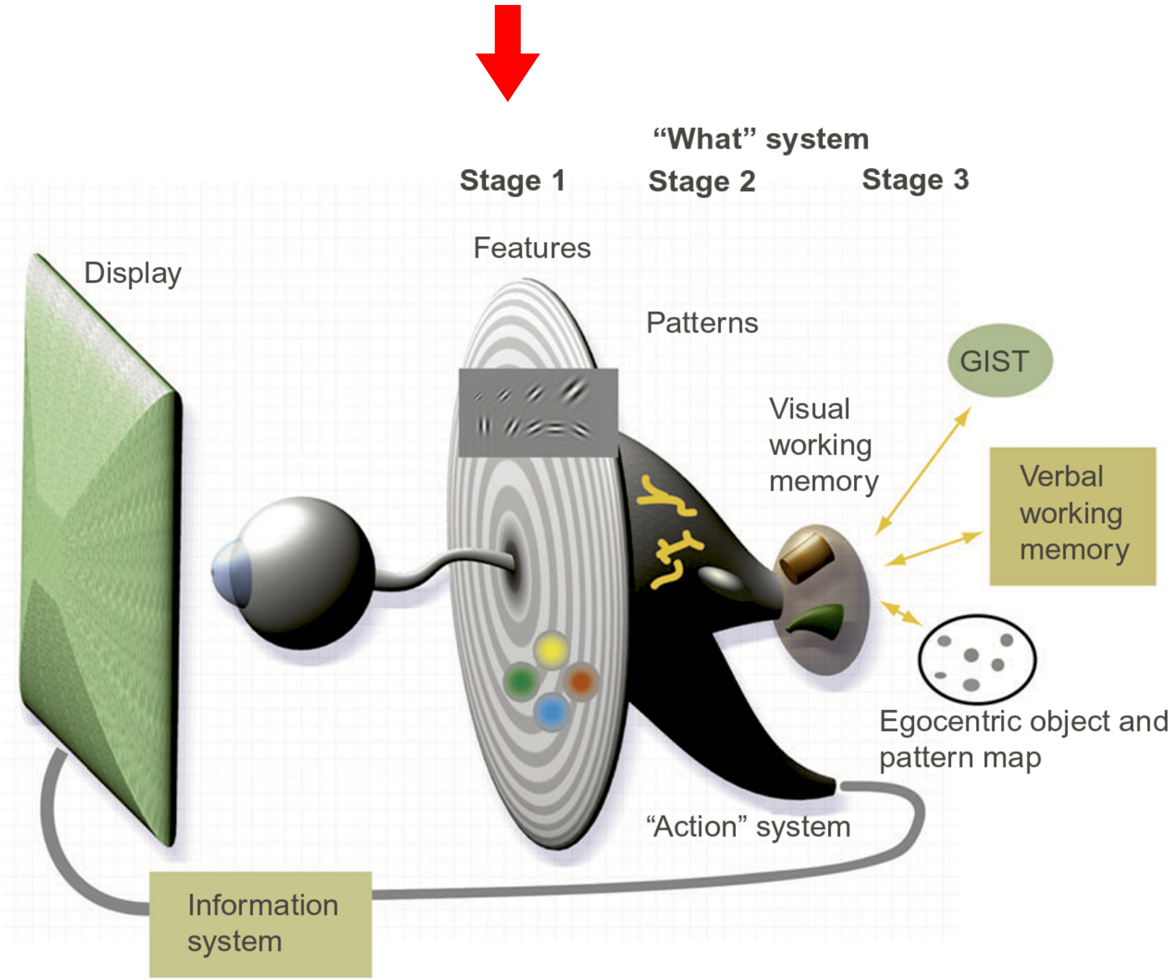


Figure 1.11 A three-stage model of visual information processing.

Source: Ware, Colin 2012.

Cognitive Scientist: Visual Information Processing

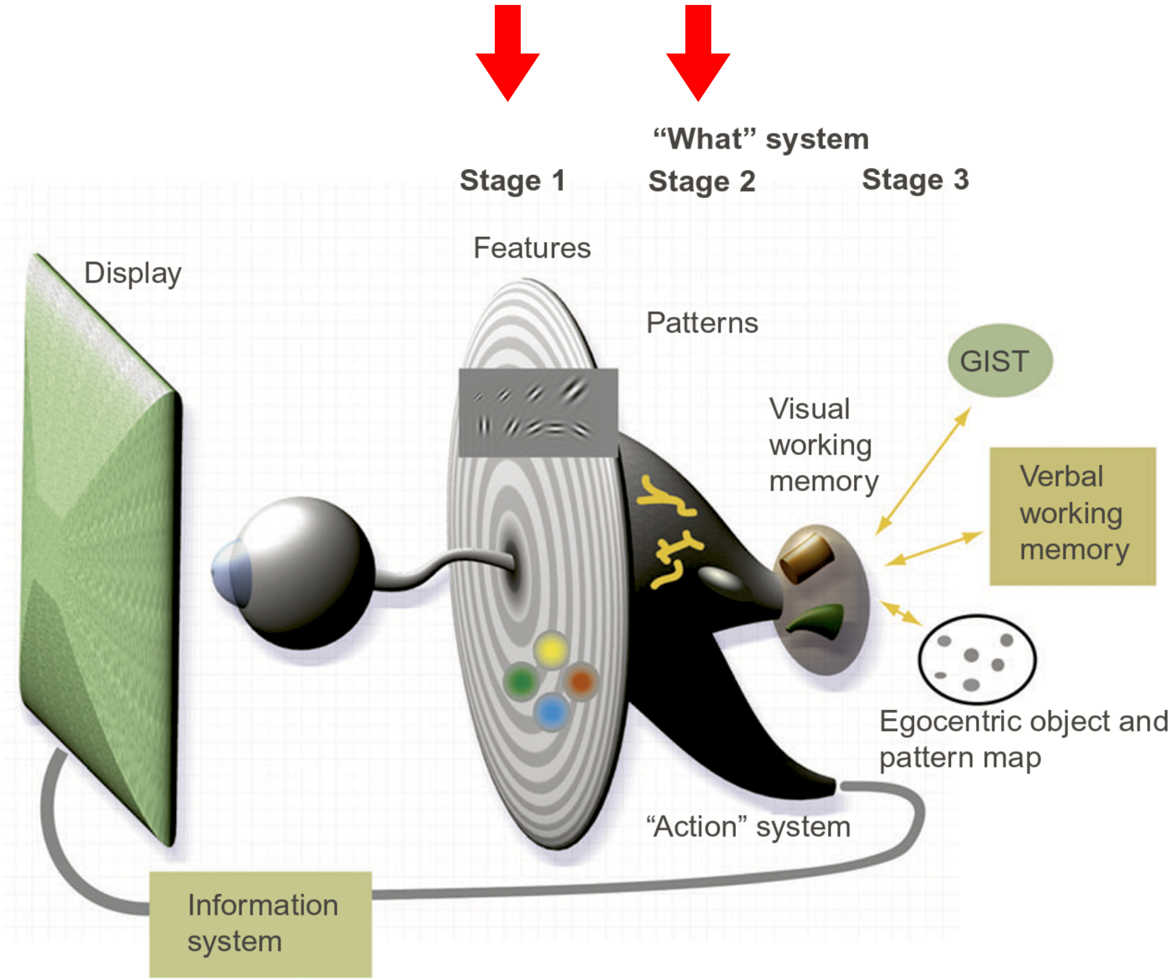


Figure 1.11 A three-stage model of visual information processing.

Source: Ware, Colin 2012.

Cognitive Scientist: Visual Information Processing

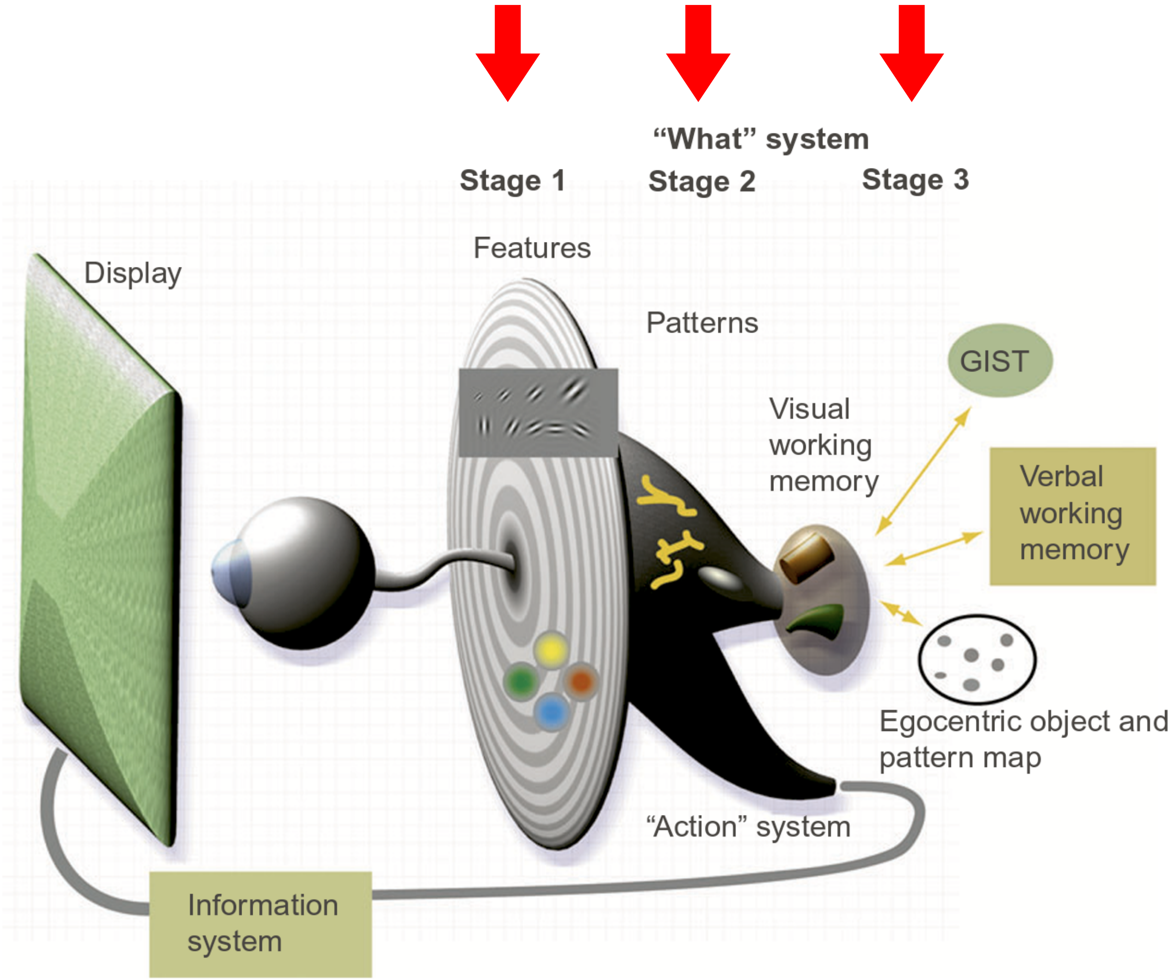


Figure 1.11 A three-stage model of visual information processing.

Source: Ware, Colin 2012.

Cognitive Scientist: Visual Information Processing

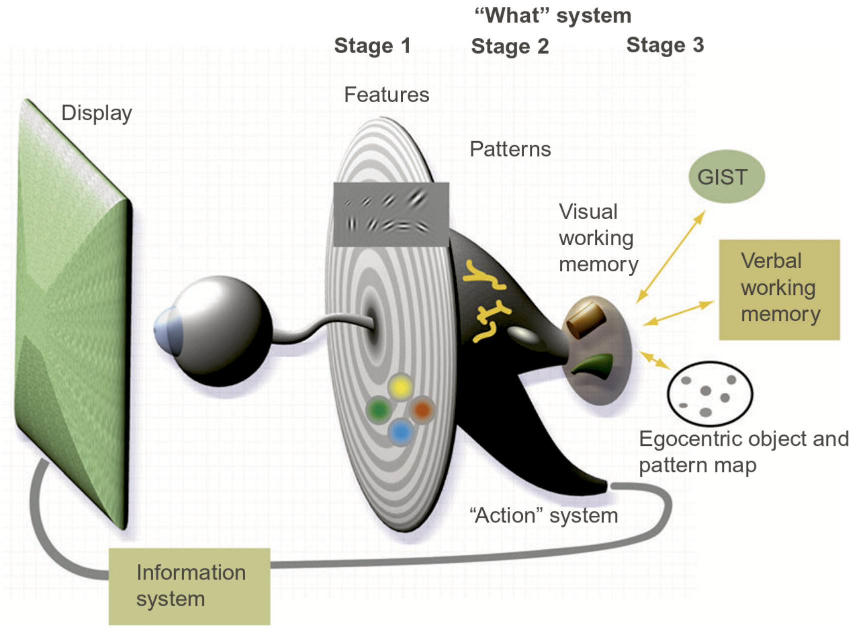


Figure 1.11 A three-stage model of visual information processing.

Source: Ware, Colin 2012.

Cognitive Scientist: Visual Information Processing

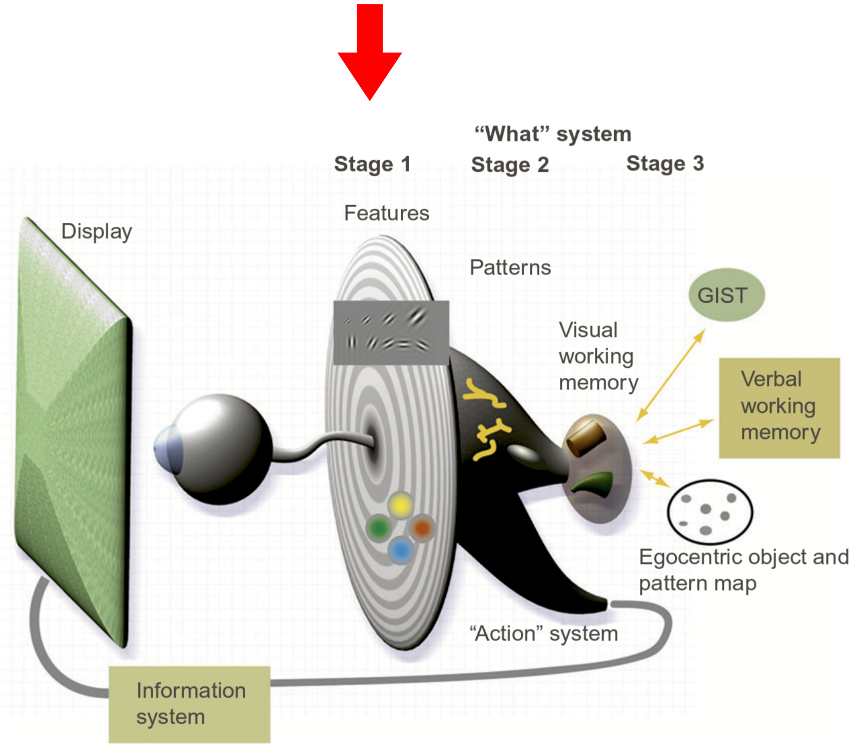


Figure 1.11 A three-stage model of visual information processing.

Source: Ware, Colin 2012.

In Stage 1, information is processed in parallel to extract basic features of the environment.

Cognitive Scientist: Visual Information Processing

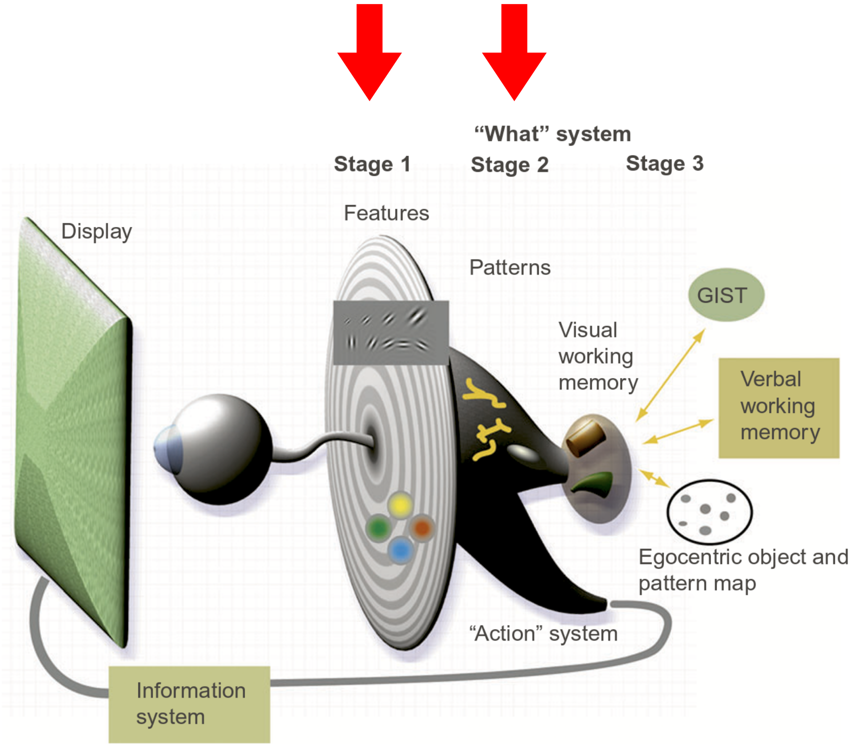


Figure 1.11 A three-stage model of visual information processing.

Source: Ware, Colin 2012.

In Stage 1, information is processed in parallel to extract basic features of the environment.

In Stage 2, active processes of pattern perception pull out structures and segment the visual scene into regions of different color, texture, and motion patterns.

Cognitive Scientist: Visual Information Processing

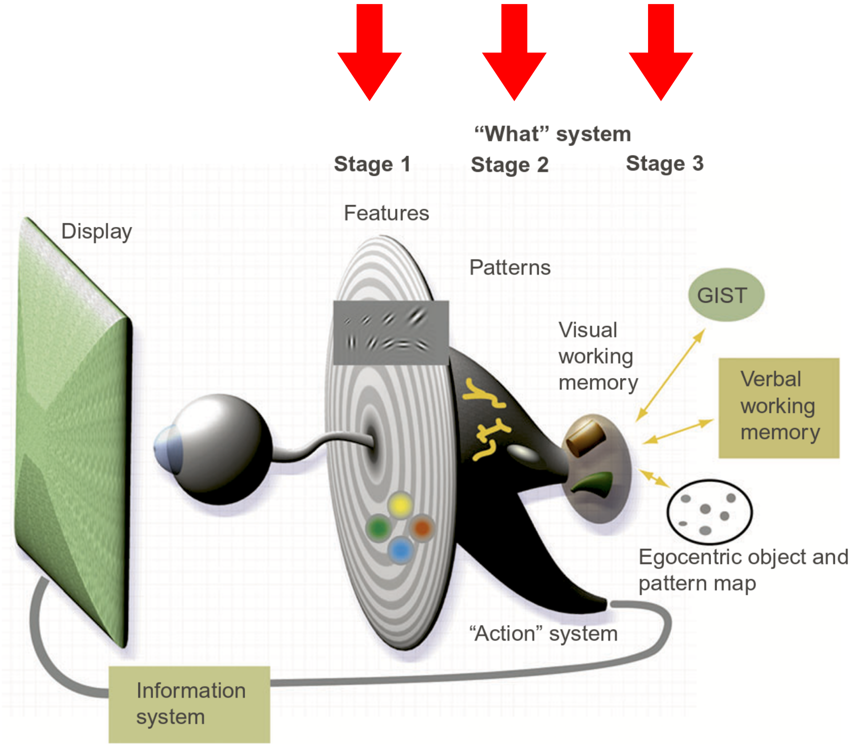


Figure 1.11 A three-stage model of visual information processing.

Source: Ware, Colin 2012.

In Stage 1, information is processed in parallel to extract basic features of the environment.

In Stage 2, active processes of pattern perception pull out structures and segment the visual scene into regions of different color, texture, and motion patterns.

In Stage 3, the information is reduced to only a few objects held in visual working memory by active mechanisms of attention to form the basis of visual thinking.

Were there no basic model of visual processing to support the idea of a good data representation, all visual representations would be arbitrary, and ultimately the problem of visualization would come down to establishing consistent notations.

Were there no basic model of visual processing to support the idea of a good data representation, all visual representations would be arbitrary, and ultimately the problem of visualization would come down to establishing consistent notations.

- Ware 2012, p.29

Cognitive Science Guidelines

Cognitive Science Guidelines

1. Design graphic representations of data by taking into account human sensory capabilities in such a way that important data elements and data patterns can be quickly perceived.
2. Important data should be represented by graphical elements that are more visually distinct than those representing less important information.

Cognitive Science Guidelines

Cognitive Science Guidelines

3. Greater numerical quantities should be represented by more distinct graphical elements.
4. Graphical symbol systems should be standardized within and across applications.
5. Where two or more tools can perform the same task, choose the one that allows for the most valuable work to be done per unit time.

Cognitive Science Guidelines

Cognitive Science Guidelines

6. Consider adopting novel design solutions only when the estimated payoff is substantially greater than the cost of learning to use them.
7. Unless the benefit of novelty outweighs the cost of inconsistency, adopt tools that are consistent with other commonly used tools.

Cognitive Science Guidelines

Cognitive Science Guidelines

8. Effort spent on developing tools should be in proportion to the profits they are expected to generate. This means that small-market custom solutions should be developed only for high-value cognitive work.

Cognitive Science

Cognitive Science

- Very important foundation for research and design of visualization of data
- Humans have very similar visual systems.
- Visual system is tuned to receive data in certain ways.
- Perceptions can be learnt.

Data Visualization: Functionalist perspective

Data Visualization: Functionalist perspective

Data Visualization User

Data Visualization: Functionalist perspective

Data Visualization User

Data Visualization Designer

Data Visualization: Functionalist perspective

Data Visualization User

Data Visualization Designer

Data Visualization Scientist

What do Cognitive Scientist teach us about visuals?

What do Cognitive Scientist teach us about visuals?

Not all graphical elements are born equal, or equally clear to our eyes.

Humans are much better at judging line length than angle or grayscale.

What do Cognitive Scientist teach us about visuals?

Not all graphical elements are born equal, or equally clear to our eyes.

Humans are much better at judging line length than angle or grayscale.

- Cleveland & McGill 1987

Graphical elements used to encode data:

1. More accurate
    1. Position on a plane
    2. Line length
    3. Angle & slope
    4. Area
    5. Volume
2. Less Accurate
    1. Color

Graphical elements used to encode data:

1. More accurate
   1. Position on a plane
   2. Line length
   3. Angle & slope
   4. Area
   5. Volume
2. Less Accurate
   1. Color

- Adoph, Cleveland & McGill 1987

Graphical elements used to encode data:

1. More accurate
    1. Position on a plane
    2. Line length
    3. Angle & slope
    4. Area
    5. Volume
2. Less Accurate
    1. Color

- Adoph, Cleveland & McGill 1987

Graphical elements used to encode data:

1. More accurate
    1. Position on a plane
    2. Line length
    3. Angle & slope
    4. Area
    5. Volume
2. Less Accurate
    1. Color Single color can stand out, …..

- Adoph, Cleveland & McGill 1987

Graphical elements used to encode data:

1. More accurate
   1. Position on a plane
   2. Line length
   3. Angle & slope
   4. Area
   5. Volume
2. Less Accurate
   1. Color

Single color can stand out, ...

- Adoph, Cleveland & McGill 1987

Graphical elements used to encode data:

1. **More accurate**
   1. **Position on a plane**
   2. **Line length**
   3. **Angle & slope**
   4. **Area**
   5. **Volume**
2. **Less Accurate**
   1. **Color**

but not multiple.

Single color can stand out, ...

- Adoph, Cleveland & McGill 1987

Graphical elements used to encode data:

1. More accurate
    1. Position on a plane
    2. Line length
    3. Angle & slope
    4. Area
    5. Volume
2. Less Accurate
    1. Color

but not multiple.

Single color can stand out, ...

- Adoph, Cleveland & McGill 1987

Graphical elements used to encode data:

Adolph and Cleveland:

Graphical elements used to encode data:

Adolph and Cleveland:

- Use more visible elements (line, points on plane) for important variables

Graphical elements used to encode data:

Adolph and Cleveland:

- Use more visible elements (line, points on plane) for important variables
- Use color for qualitative data (e.g. countries)

Angular data encoding

Angular data encoding

Which band

is

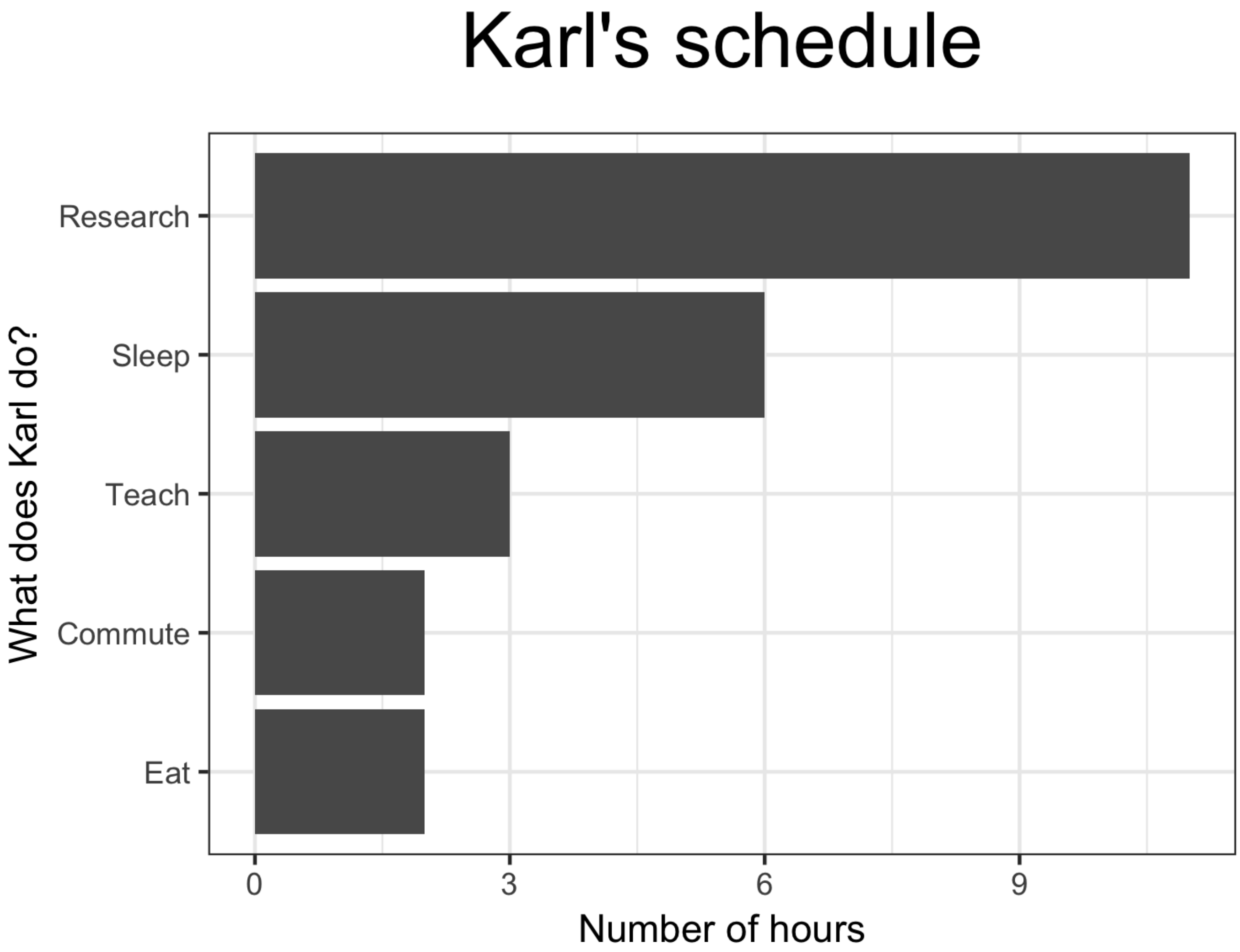
what?

Angular data encoding

Which band
is
what?

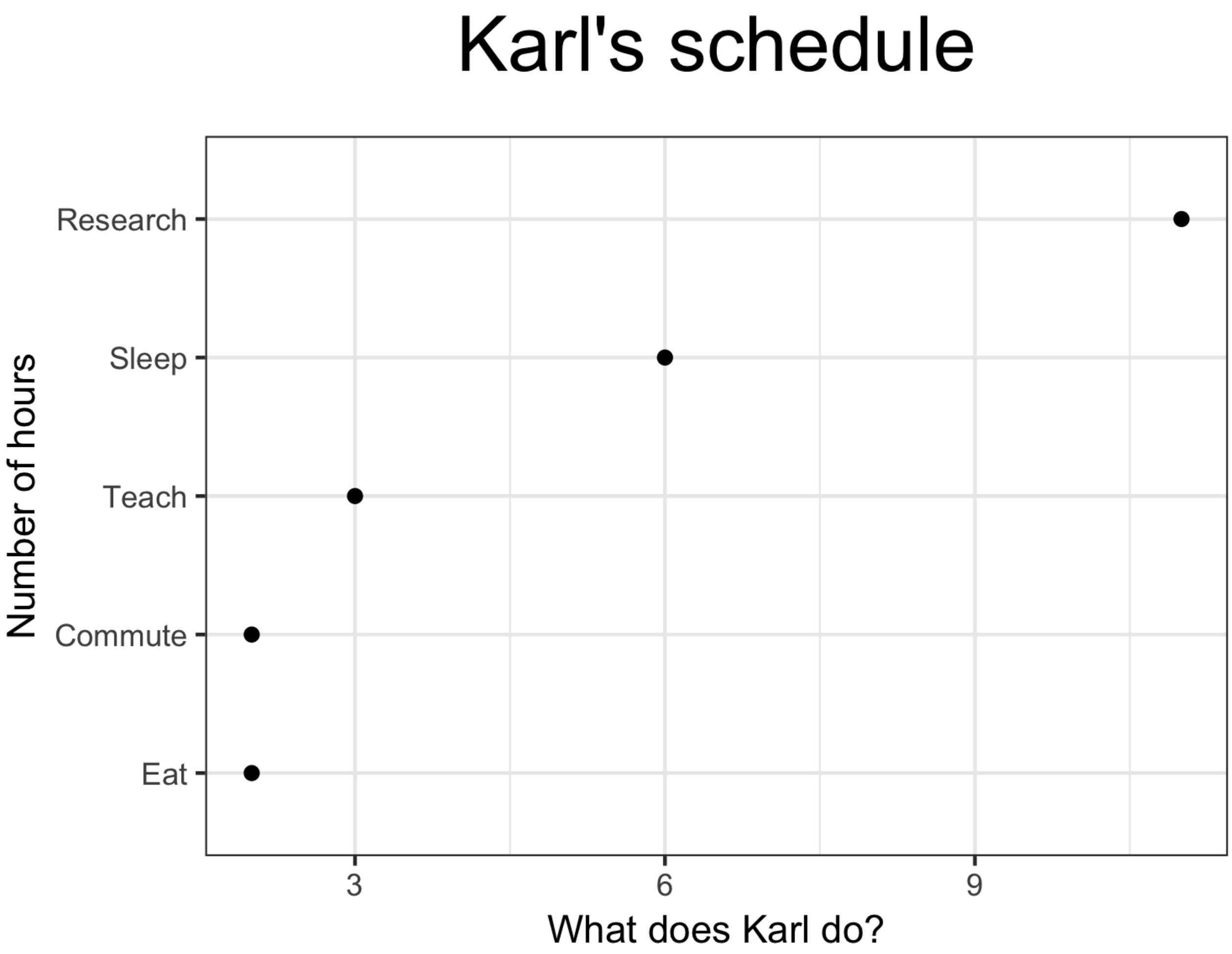
Source: Google charts based on javascript

Better



Karl's schedule
Research
Sleep
Teach
Commute
Eat
What does Karl do?
0
3
6
9
Number of hours

Better better



Find the 3’s

85689726984689762689764358922659865986554897689269898

02462996874026557627986789045679232769285460986772098

90834579802790759047098279085790847729087590827908754

98709856749068975786259845690243790472190790709811450

85689726984689762689764458922659865986554897689269898

Find the 3's

85689726984689762689764358922659865986554897689269898

02462996874026557627986789045679232769285460986772098

90834579802790759047098279085790847729087590827908754

98709856749068975786259845690243790472190790709811450

85689726984689762689764458922659865986554897689269898

Find the 3’s

8568972698468976268976435892265986598655489768926989 8

0246299687402655762798678904567923276928546098677209 8

9083457980279075904709827908579084772908759082790875 4

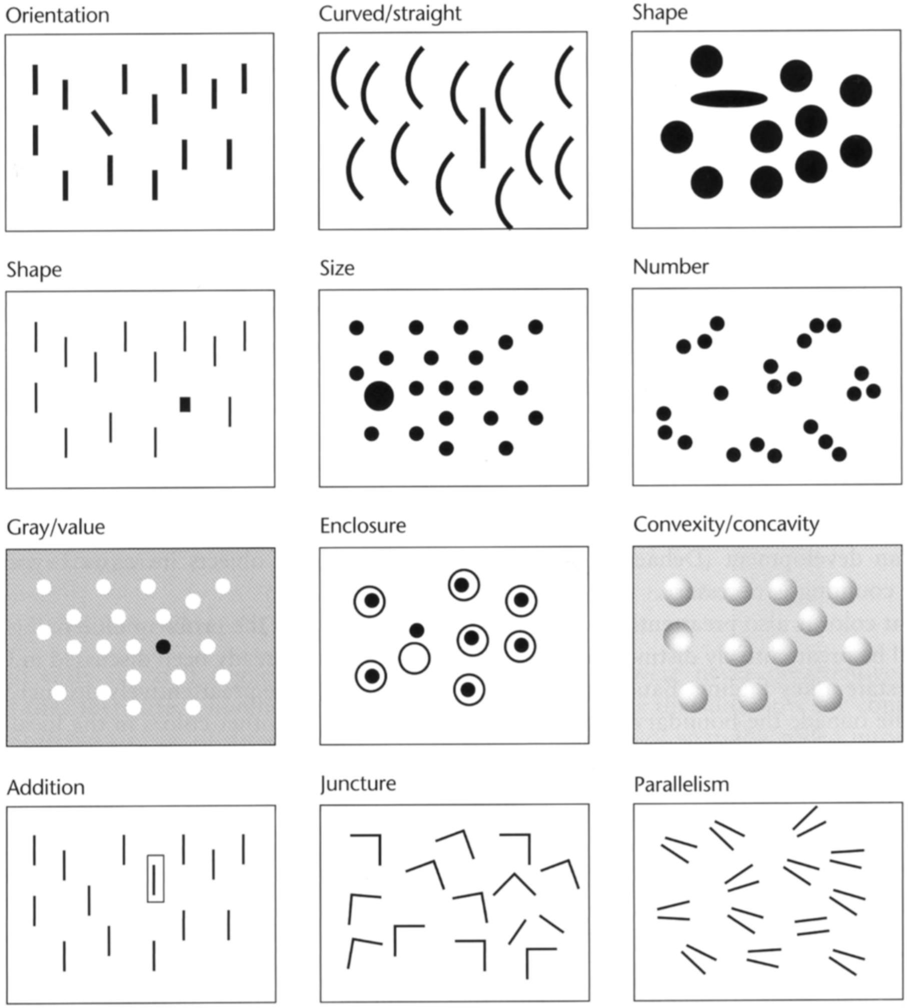
9870985674906897578625984569024379047219079070981145 0

8568972698468976268976445892265986598655489768926989 8

Our brains process color differences “pre-attentively” – fast & effortlessly.

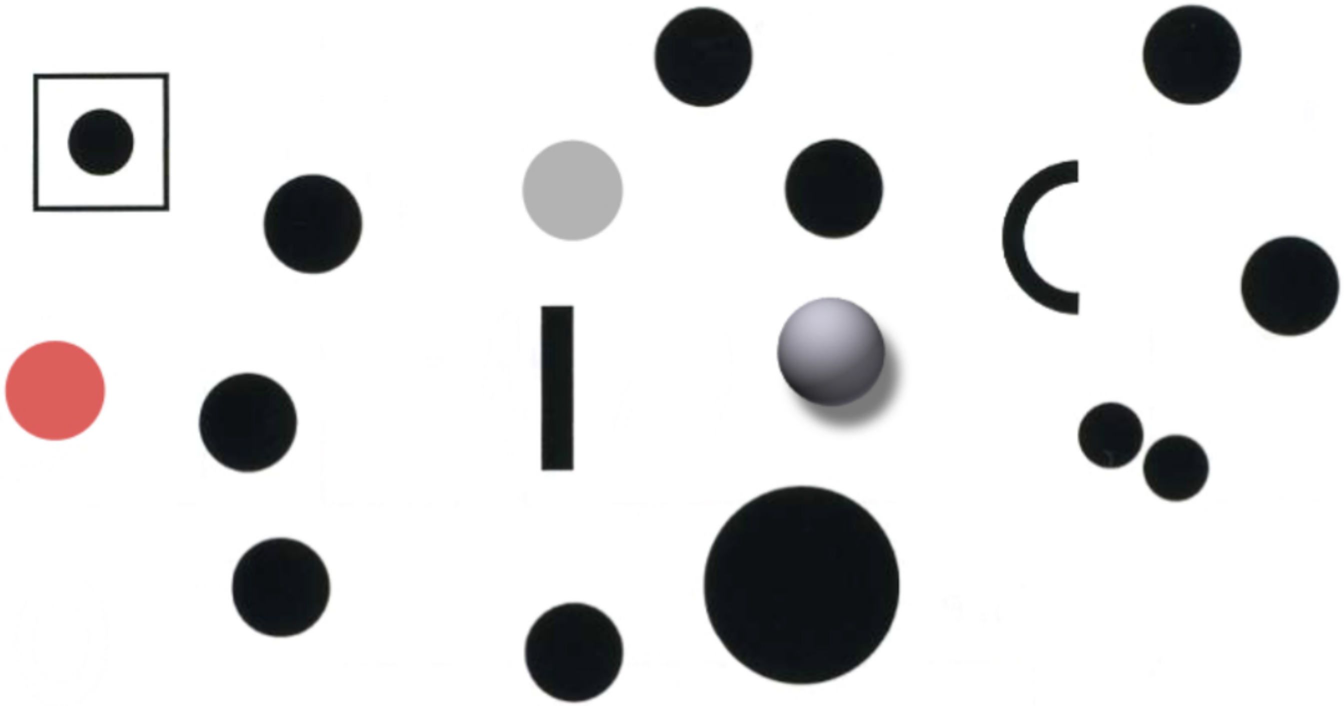
- Colin Ware

Shapes too


Orientation
Curved/straight
Shape
Shape
Size
Number
Gray/value
Enclosure
Convexity/concavity
Addition
Juncture
Parallelism

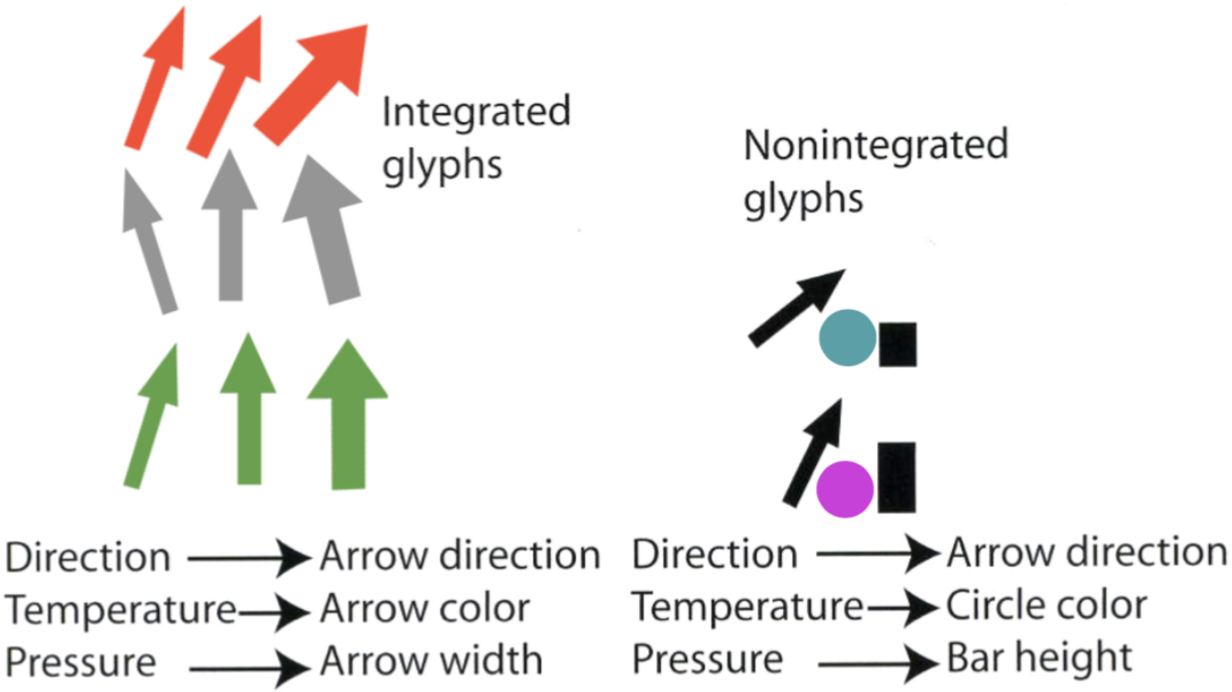
Again, not all are equal...

Again, not all are equal...

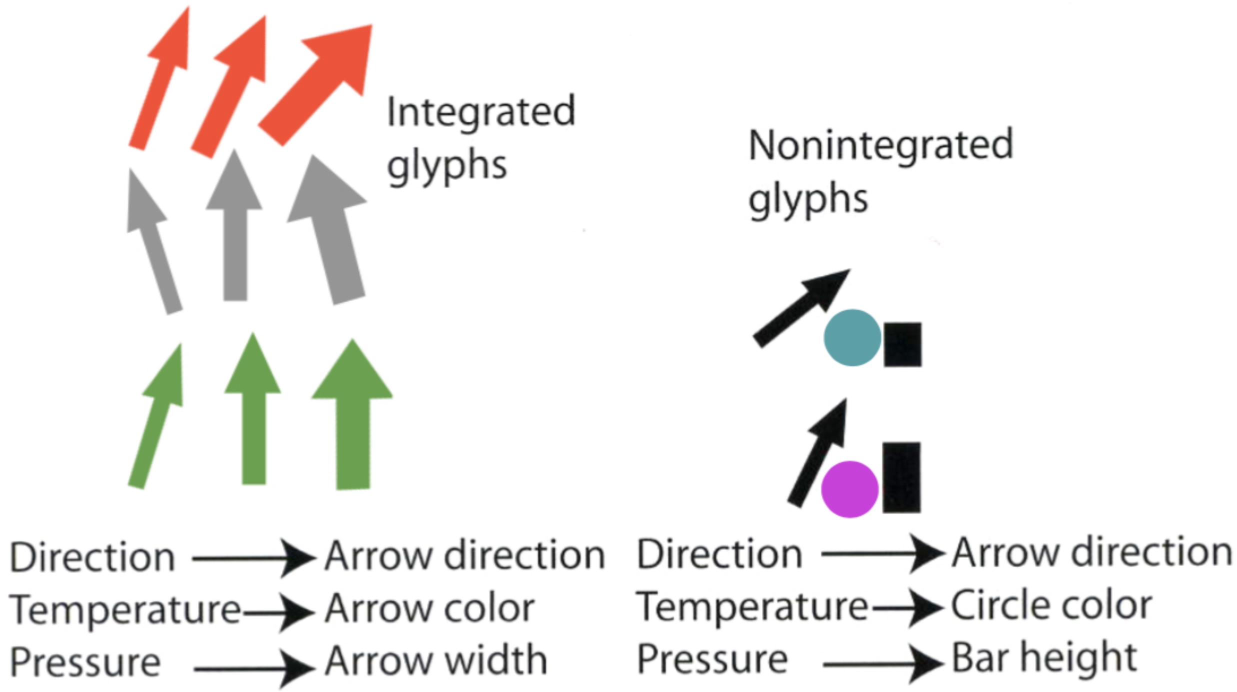


Multidimensional glyth

Multidimensional glyth



Multidimensional glyth



The more variables you encode to dimensions of glyphs, the harder it is to pre-attentively separate the dimensions.

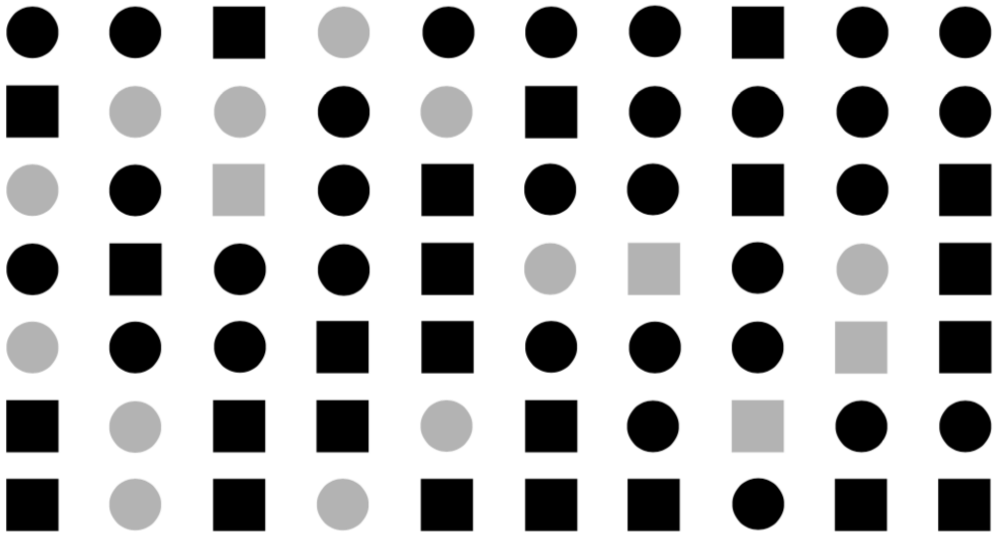
Multidimensional glyth

Multidimensional glyth

Caveat: The more variables you encode to dimensions of glyphs, the harder it is to pre-attentively separate the dimensions

Multidimensional glyth

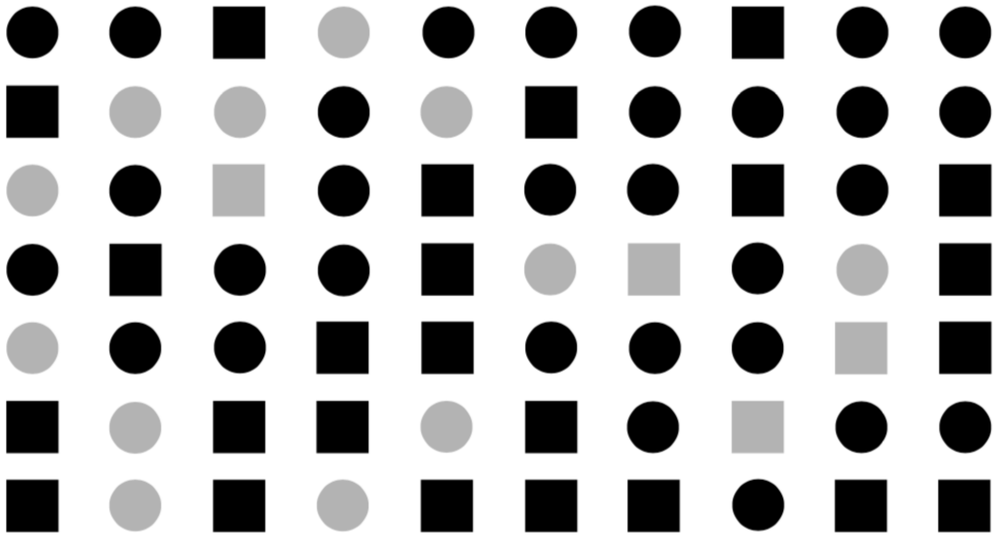
Caveat: The more variables you encode to dimensions of glyphs, the harder it is to pre-attentively separate the dimensions



Multidimensional glyth

Caveat: The more variables you encode to dimensions of glyphs, the harder it is to pre-attentively separate the dimensions

How many gray circles?



Color

Color

1. Choose colors for quantities using pre-attentively smooth gradients

Color

1. Choose colors for quantities using pre-attentively smooth gradients
2. Choose colors for categories to achieve equal pairwise distinctions

Color

1. Choose colors for quantities using pre-attentively smooth gradients
2. Choose colors for categories to achieve equal pairwise distinctions
3. Avoid overlapping colors with similar brightness (value)

Color

1. Choose colors for quantities using pre-attentively smooth gradients
2. Choose colors for categories to achieve equal pairwise distinctions
3. Avoid overlapping colors with similar brightness (value)
4. Use pastels for large area colors and saturated colors for small points

Color

1. Choose colors for quantities using pre-attentively smooth gradients
2. Choose colors for categories to achieve equal pairwise distinctions
3. Avoid overlapping colors with similar brightness (value)
4. Use pastels for large area colors and saturated colors for small points

- Chris Adoph

Color Science

Color Science

1. a measure of the wavelength of light

Color Science

1. a measure of the wavelength of light
2. Measured by
   1. a mixture of red, blue, and green
   2. a mixture of hue, luminosity, and saturation

Color Science

1. a measure of the wavelength of light
2. Measured by
   1. a mixture of red, blue, and green
   2. a mixture of hue, luminosity, and saturation
3. Categorical but also could be continuous if perceivable

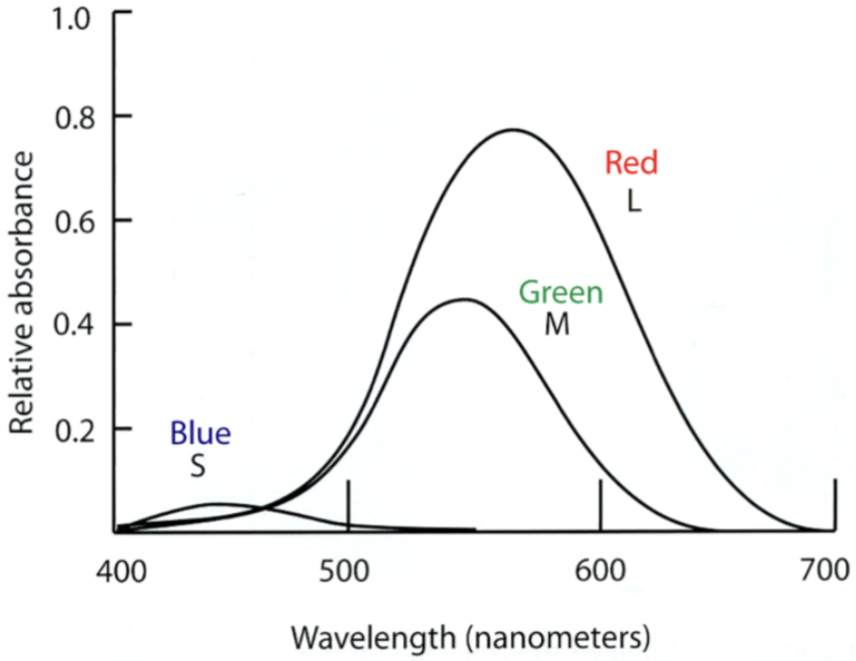
Color Science

1. a measure of the wavelength of light
2. Measured by
   1. a mixture of red, blue, and green
   2. a mixture of hue, luminosity, and saturation
3. Categorical but also could be continuous if perceivable

- Chris Adoph

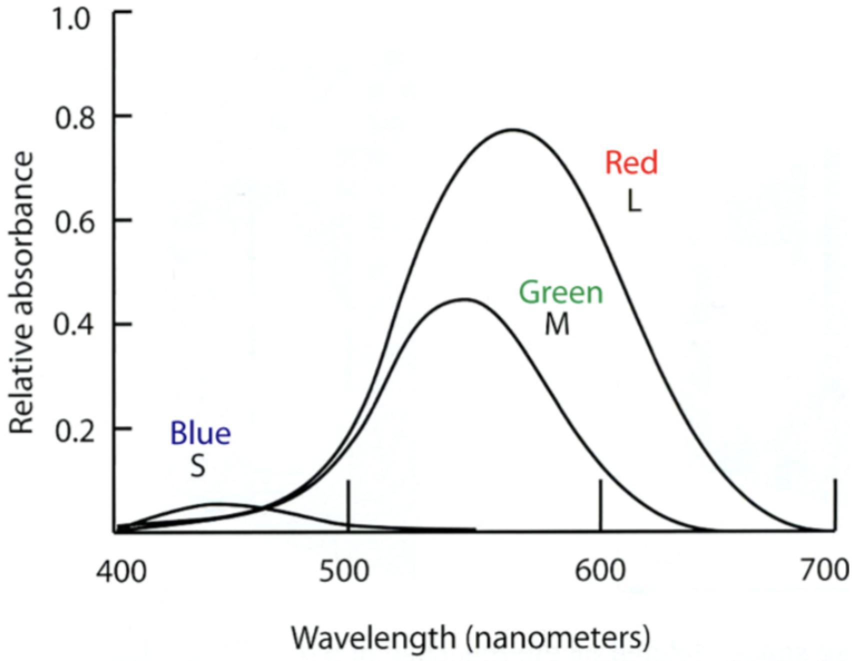
Color Science

Color Science

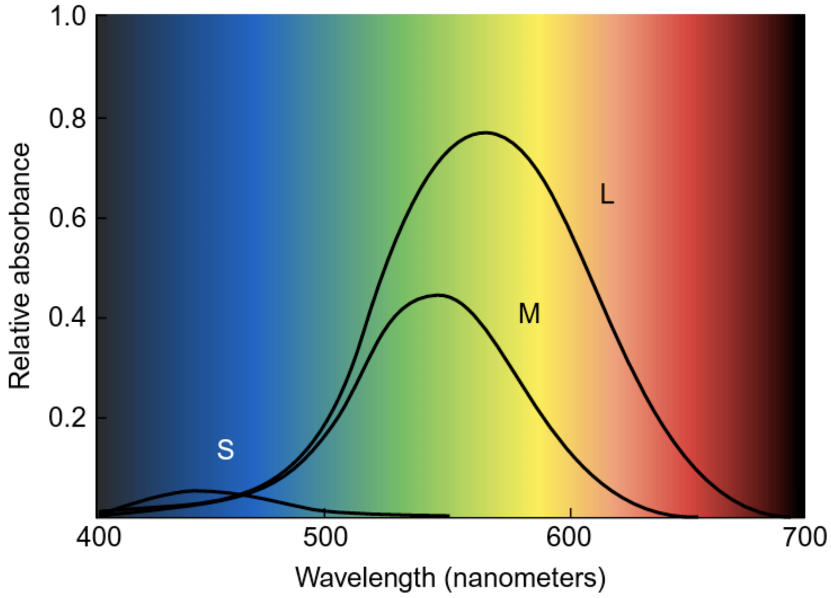


- Chris Adoph

Color Science



- Chris Adoph



- Colin Ware

Color Science

Color Science

We can describe a color by the following equation:

Color Science

We can describe a color by the following equation:

$$C = rR + gG + bB$$

Color Science

We can describe a color by the following equation:

$$C = rR + gG + bB$$

where C is the color to be matched; R , G , and B are the primary light sources to be used to create a match; and r , g , and b represent the amounts of each primary light.

Color Science

Color Science

Human eyes contain two kinds of photo-receptive elements:

1. Rods
   - Sensitive to brightness
   - Single photon receptors
   - Little use in sunlight
2. Cones
   - Come in three varieties...
   1. Short wavelength (red) - most sensitive
   2. Medium wavelength (green) - moderately sensitive
   3. Long wavelength (blue) - weak

Color Science

Humans are:

Color Science

Humans are:

- best at seeing red

Color Science

Humans are:

- best at seeing red
- worst at seeing blue

Color Science

Humans are:

- best at seeing red
- worst at seeing blue

Species vary in color vision ability:

- dogs have only two cones, are red-green colorblind, and see less detail in daylight
- birds have more cones than humans – chickens have 12!

Color Science

Humans are:

- best at seeing red
- worst at seeing blue

Species vary in color vision ability:

- dogs have only two cones, are red-green colorblind, and see less detail in daylight
- birds have more cones than humans – chickens have 12!

Number of cones = number of primary colors a species perceives.
Mixing the three (human) primaries in different amounts makes any color humans can see.

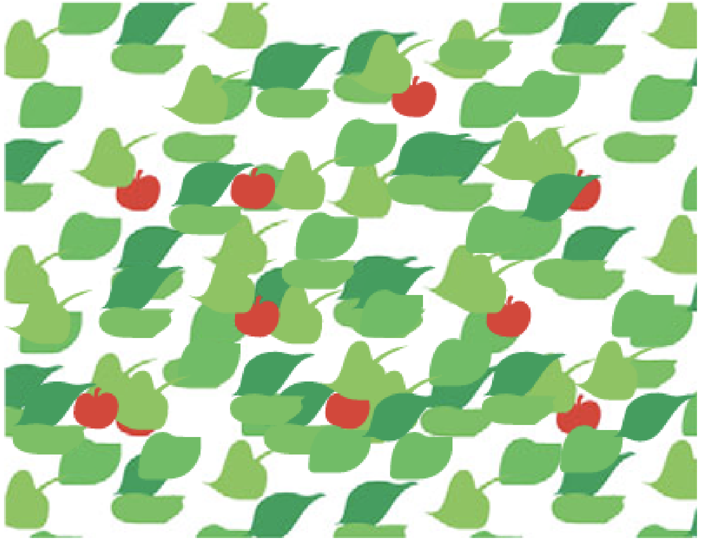
Color Blindness

Color Blindness

Color Blindness: the most common deficiencies are explained by lack of either the long-wavelength-sensitive cones (protanopia) or the medium-wavelength-sensitive cones (deuteranopia). Both protanopia and deuteranopia result in an inability to distinguish red and green, meaning that the cherries below are difficult for people with these deficiencies to see.

Color Blindness

Color Blindness: the most common deficiencies are explained by lack of either the long-wavelength-sensitive cones (protanopia) or the medium-wavelength-sensitive cones (deuteranopia). Both protanopia and deuteranopia result in an inability to distinguish red and green, meaning that the cherries below are difficult for people with these deficiencies to see.



Color Science

Primaries and color can be expressed in many equivalent ways.

These are different colorspaces: mappings from 3 variables to a color:

Computer space · RGB Red, Green, Blue
Printer space · CMYK Cyan, Magenta, Yellow, Black
Artist space · HSV
Hue, Saturation, Value
Brain space · CIElab
Lightness, blue/yellow, red/green

Color Spaces

Color Spaces

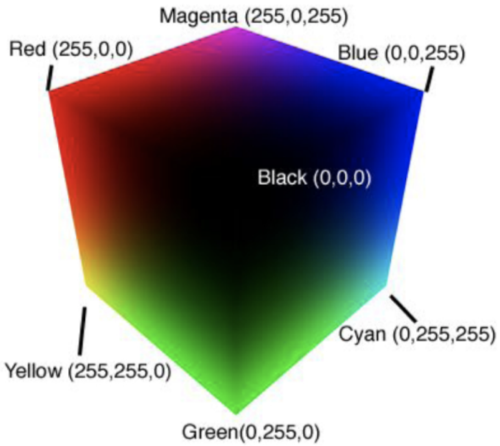


Figure 3a: The black corner (0,0,0) of the RGB Cube

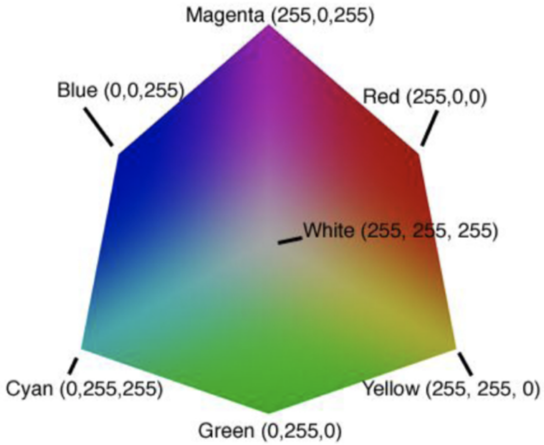


Figure 3b: The white corner (255,255,255) of the RGB Cube

- Chris Adolph

Color Spaces: Artists

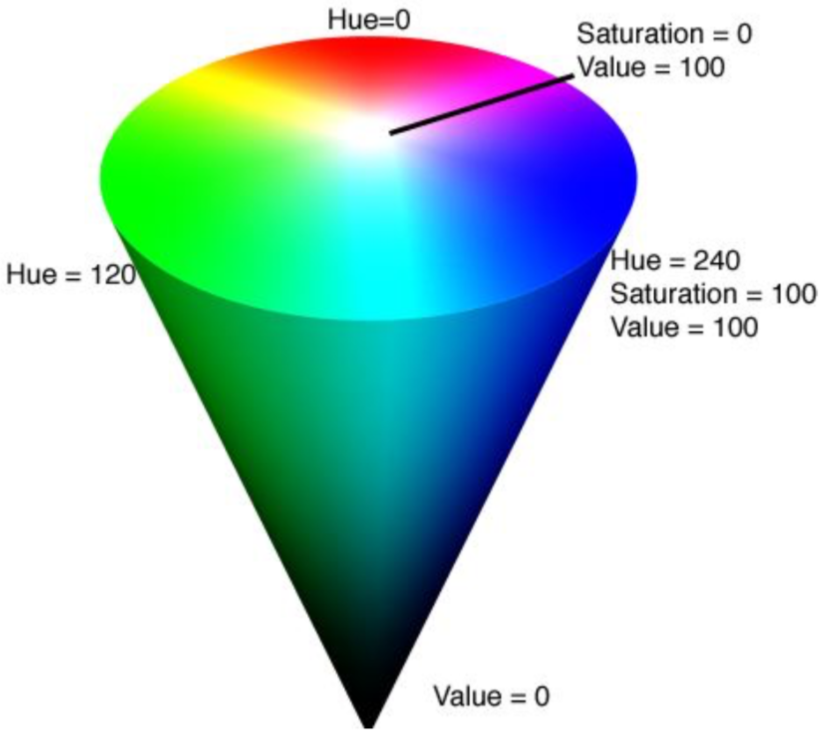


Figure 2: The HSV Cone

Hue is the "name" of the color

Saturation is the richness of the color; desaturated colors have been mixed with gray

Value is the brightness of the color

Color Spaces: Artists

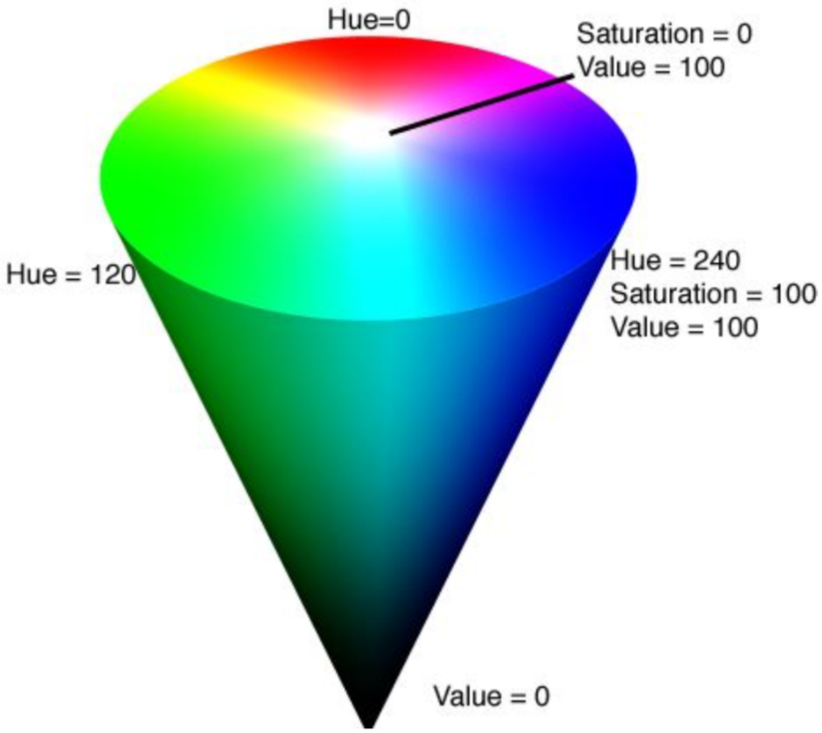


Figure 2: The HSV Cone

Hue is the "name" of the color

Saturation is the richness of the color; desaturated colors have been mixed with gray

Value is the brightness of the color

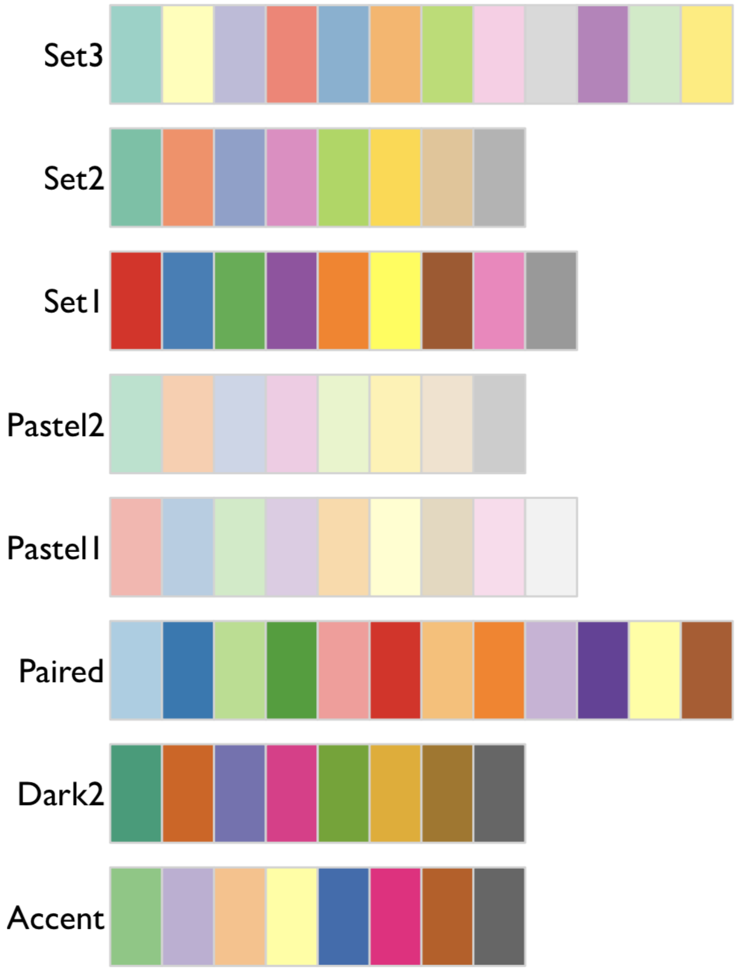
Source: Darrin Cardani, "Adventures in the HSV Space"

Color in Data Visualization

in R:

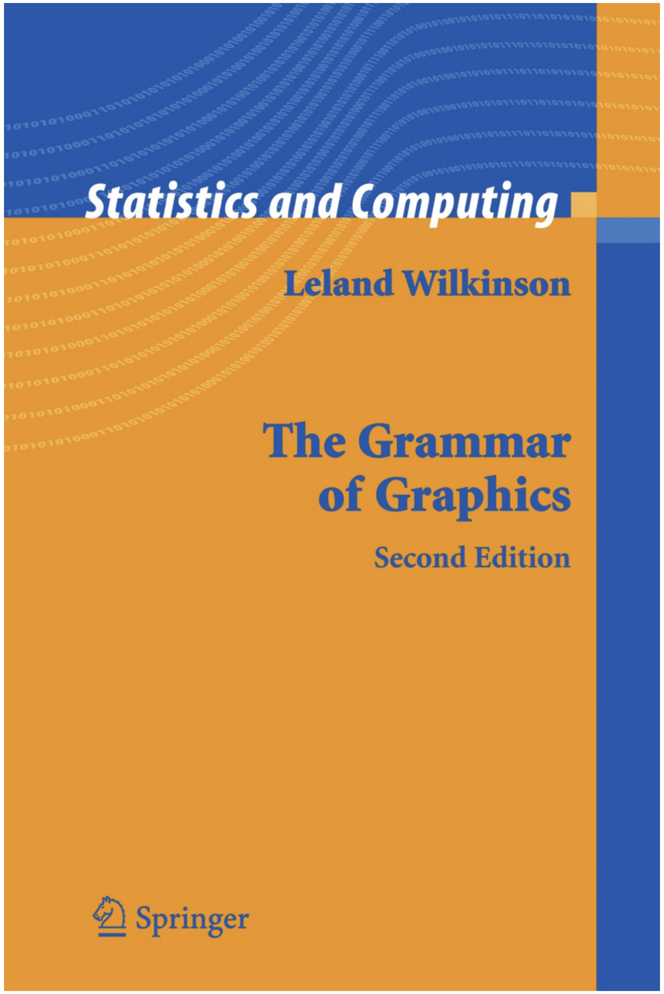
RColorBrewer will choose appropriate colors for you

Color in Data Visualization

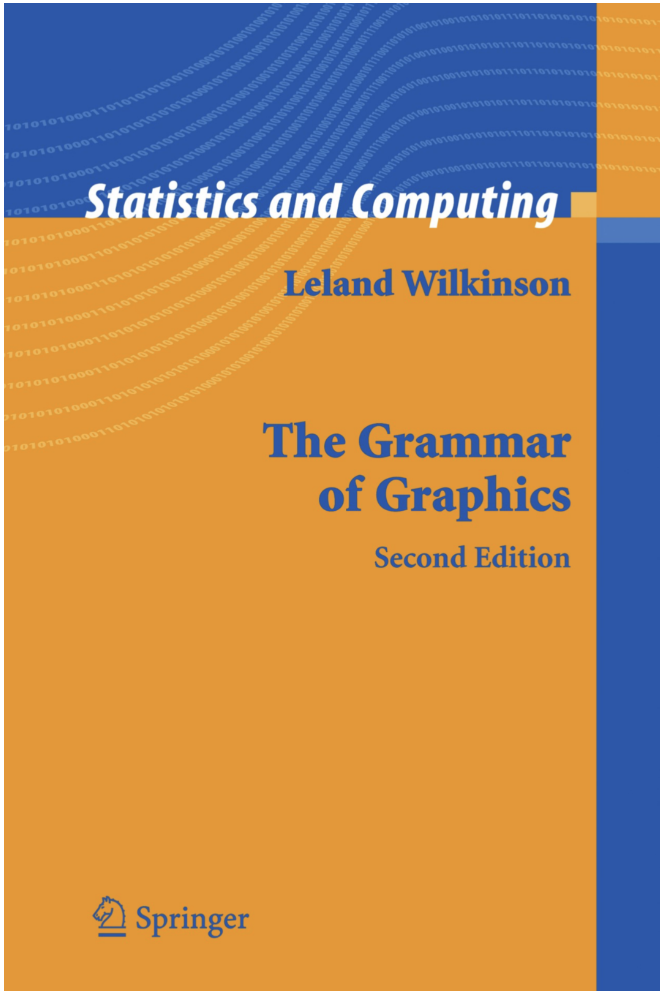


library(RColorBrewer)

display.brewer.all(type="qual")

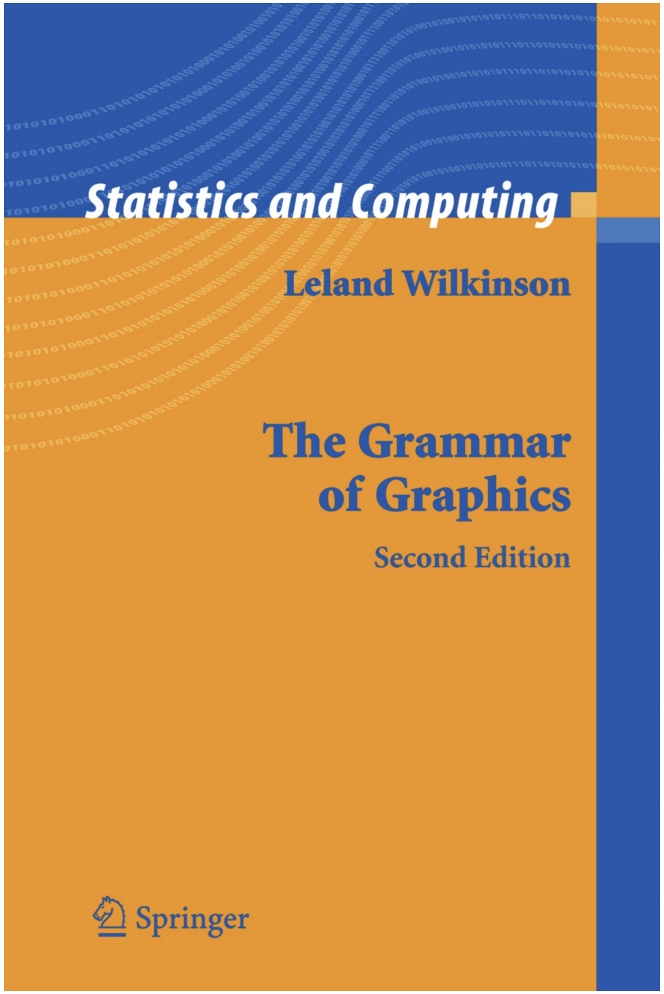

Statistics and Computing
Leland Wilkinson
The Grammar of Graphics
Second Edition
Springer

ggplot2


Statistics and Computing
Leland Wilkinson
The Grammar of Graphics
Second Edition
Springer

ggplot2

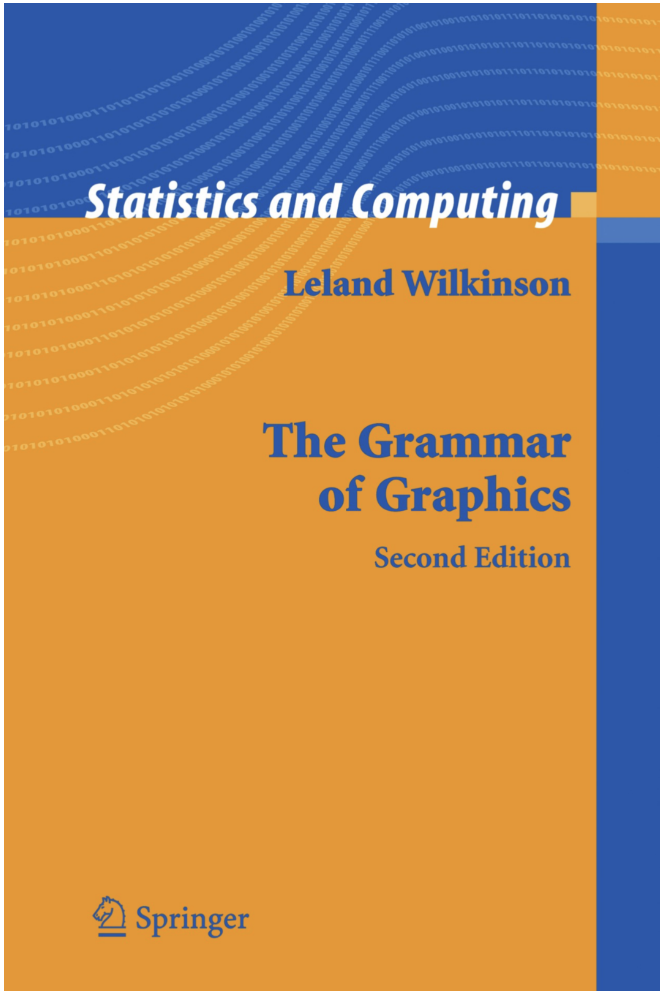
Author: Hadley Wickham (from New Zealand, *Rstudio* Chief Scientist)



ggplot2

Author: Hadley Wickham (from New Zealand, *Rstudio* Chief Scientist)

Based on based Leland Wilkinson's book, *The Grammar of Graphics* (2005, available online at UTD library)



Wickham on ggplot2

Wickham on ggplot2

Aim of grammar:

“bring together in a coherent way things that previously appeared unrelated and which also will provide a basis for dealing systematically with new situations”.

- D. R. Cox 1978

Wilkinson

Wilkinson

Grammar gives language rules.

Wilkinson

Grammar gives language rules.

The word stems from the Greek noun for letter or mark $\gamma\rho\acute{\alpha}\mu\mu\alpha$.

Wilkinson

Grammar gives language rules.

The word stems from the Greek noun for letter or mark $\gamma\rho\acute{\alpha}\mu\mu\alpha$.

That derives from the Greek verb for writing , $\gamma\rho\acute{\alpha}\phi\omega$ is the source of our English word ***graph.***

Wilkinson

Wilkinson

A grammar is a formal system of rules for generating lawful statements in a language.

Wilkinson

A grammar is a formal system of rules for generating lawful statements in a language.

The grammar of graphics goes beyond a limited set of charts (words) to an unlimited world of graphical forms (statements).

Wilkinson

A grammar is a formal system of rules for generating lawful statements in a language.

The grammar of graphics goes beyond a limited set of charts (words) to an unlimited world of graphical forms (statements).

The rules of graphics grammar are sometimes mathematical and sometimes aesthetic.

Wilkinson

Wilkinson

Mathematics provides symbolic tools for representing abstractions.

Wilkinson

Mathematics provides symbolic tools for representing abstractions.

Aesthetics, in the original Greek sense, offers principles for relating sensory attributes (color, shape, sound, etc.) to abstractions..

Wilkinson

Mathematics provides symbolic tools for representing abstractions.

Aesthetics, in the original Greek sense, offers principles for relating sensory attributes (color, shape, sound, etc.) to abstractions..

GG focuses on rules for constructing graphs mathematically and then representing them as graphics aesthetically.

Wickham

Wickham

A grammar of graphics is a tool that enables concise description of the components of a graphic.

Such a grammar allows moving beyond named graphics and gain insight into the deep structure that underlies statistical graphics.

Wickham on ggplot2

Wickham on ggplot2

"layered grammar of graphics"

1. develop a hierarchy of defaults based on Wilkinson
2. embed a graphical grammar into a programming language.
3. build on the grammar to learn how to create graphical “poems.”

ggplot2

ggplot2

Two ways of plotting a graphic:

ggplot2

Two ways of plotting a graphic:

1. qplot() - quick plot

ggplot2

Two ways of plotting a graphic:

1. qplot() - quick plot
2. ggplot() - grammar of graphics plot

ggplot2

ggplot2

Two ways of plotting a graphic:

1. qplot() - quick plot

ggplot2

Two ways of plotting a graphic:

1. qplot() - quick plot
    1. very similar to plot()

ggplot2

Two ways of plotting a graphic:

1. qplot() - quick plot
    1. very similar to plot()
    2. simple to use

ggplot2

Two ways of plotting a graphic:

1. qplot() - quick plot
    1. very similar to plot()
    2. simple to use
    3. quick to produce basic graphs

ggplot2

Two ways of plotting a graphic:

ggplot2

Two ways of plotting a graphic:

2. ggplot() - grammar of graphics plot

ggplot2

Two ways of plotting a graphic:

2. ggplot() - grammar of graphics plot
    1. fuller implementation of The Grammar of Graphics

ggplot2

Two ways of plotting a graphic:

2. ggplot() - grammar of graphics plot
    1. fuller implementation of The Grammar of Graphics
    2. highly flexible for plotting graphs

ggplot2

Two ways of plotting a graphic:

2. ggplot() - grammar of graphics plot
    1. fuller implementation of The Grammar of Graphics
    2. highly flexible for plotting graphs
    3. steep learning curve

ggplot2 components

ggplot2 components

Create graphical display by combining building blocks including:

ggplot2 components

Create graphical display by combining building blocks including:

- data

ggplot2 components

Create graphical display by combining building blocks including:

- data
- aesthetic mapping

ggplot2 components

Create graphical display by combining building blocks including:

- data
- aesthetic mapping
- geometric object

ggplot2 components

Create graphical display by combining building blocks including:

- data
- aesthetic mapping
- geometric object
- statistical transformations

ggplot2 components

Create graphical display by combining building blocks including:

- data
- aesthetic mapping
- geometric object
- statistical transformations
- scales

ggplot2 components

Create graphical display by combining building blocks including:

- data
- aesthetic mapping
- geometric object
- statistical transformations
- scales
- coordinate system

ggplot2 components

Create graphical display by combining building blocks including:

- data
- aesthetic mapping
- geometric object
- statistical transformations
- scales
- coordinate system
- position adjustments

ggplot2 components

Create graphical display by combining building blocks including:

- data
- aesthetic mapping
- geometric object
- statistical transformations
- scales
- coordinate system
- position adjustments
- faceting

Components of ggplot2

Components of ggplot2

1. **data**: R data frame

Components of ggplot2

1. **data**: R data frame
2. **coordinate syste**m: 2-D space plot

Components of ggplot2

1. **data**: R data frame
2. **coordinate syste**m: 2-D space plot
3. **geoms**: geometric objects representing data, e.g. points, lines, polygons, etc.

Components of ggplot2

1. **data**: R data frame
2. **coordinate syste**m: 2-D space plot
3. **geoms**: geometric objects representing data, e.g. points, lines, polygons, etc.
4. **aesthetics**: visual characteristics, e.g. position, size, color, shape, transparency, fill

Components of ggplot2

1. **data**: R data frame
2. **coordinate syste**m: 2-D space plot
3. **geoms**: geometric objects representing data, e.g. points, lines, polygons, etc.
4. **aesthetics**: visual characteristics, e.g. position, size, color, shape, transparency, fill
5. **scales**: governs how visual characteristic is converted to display values, e.g. log scales, color scales, size scales, shape scales.

Components of ggplot2

1. **data**: R data frame
2. **coordinate syste**m: 2-D space plot
3. **geoms**: geometric objects representing data, e.g. points, lines, polygons, etc.
4. **aesthetics**: visual characteristics, e.g. position, size, color, shape, transparency, fill
5. **scales**: governs how visual characteristic is converted to display values, e.g. log scales, color scales, size scales, shape scales.
6. **stats**: statistical data transformations, e.g. counts, means, medians, regression lines

Components of ggplot2

1. **data**: R data frame
2. **coordinate syste**m: 2-D space plot
3. **geoms**: geometric objects representing data, e.g. points, lines, polygons, etc.
4. **aesthetics**: visual characteristics, e.g. position, size, color, shape, transparency, fill
5. **scales**: governs how visual characteristic is converted to display values, e.g. log scales, color scales, size scales, shape scales.
6. **stats**: statistical data transformations, e.g. counts, means, medians, regression lines
7. **facets**: split data into subsets to display as multiple graphs

ggplot2

Philosophy:

ggplot2

Philosophy:

- Present data with layers of components

ggplot2

Philosophy:

- Present data with layers of components
- tidy data concepts (tidyverse.org)

ggplot2

Philosophy:

- Present data with layers of components
- tidy data concepts (tidyverse.org)
- Grammar structure enables graphical data analysis and exploration

Gapminder data set

The gapminder data set has 1704 rows and 6 variables:

country(factor) - 142 levels

continent(factor) - 5 levels

year -ranges from 1952 to 2007 in increments of 5 years

lifeExp- life expectancy at birth, in years

pop - population

gdpPercap - GDP per capita (US$, inflation-adjusted)

Gapminder data set

```
install.packages("gapminder")
library(gapminder)
gm=gapminder
head(gm)
summary(gm)
table(gm$country)
```

plot() in graphics package

```
# Plot one variable
hist(gm$lifeExp)
```

```
# Plot two variables with logged version of x
plot(lifeExp ~ gdpPercap, gm, subset = year == 2007, log =
"x", pch=16)
```

Plot type

Plot type

- "p" for points

Plot type

- "p" for points
- "l" for lines

Plot type

- "p" for points
- "l" for lines
- "b" for both

Plot type

- "p" for points
- "l" for lines
- "b" for both
- "c" for the lines part alone of "b"

Plot type

- "p" for points
- "l" for lines
- "b" for both
- "c" for the lines part alone of "b"
- "o" for both ‘overplotted’

Plot type

- "p" for points
- "l" for lines
- "b" for both
- "c" for the lines part alone of "b"
- "o" for both ‘overplotted’
- "h" for ‘histogram’ like (or ‘high-density’) vertical lines

Plot type

- "p" for points
- "l" for lines
- "b" for both
- "c" for the lines part alone of "b"
- "o" for both ‘overplotted’
- "h" for ‘histogram’ like (or ‘high-density’) vertical lines
- "s" for stair steps, moves first horizontal, then vertical

Plot type

- "p" for points
- "l" for lines
- "b" for both
- "c" for the lines part alone of "b"
- "o" for both ‘overplotted’
- "h" for ‘histogram’ like (or ‘high-density’) vertical lines
- "s" for stair steps, moves first horizontal, then vertical
- "S" for other steps, contrary to "s"

Plot type

- "p" for points
- "l" for lines
- "b" for both
- "c" for the lines part alone of "b"
- "o" for both ‘overplotted’
- "h" for ‘histogram’ like (or ‘high-density’) vertical lines
- "s" for stair steps, moves first horizontal, then vertical
- "S" for other steps, contrary to "s"
- "n" for no plotting.

Plot type

- "p" for points
- "l" for lines
- "b" for both
- "c" for the lines part alone of "b"
- "o" for both ‘overplotted’
- "h" for ‘histogram’ like (or ‘high-density’) vertical lines
- "s" for stair steps, moves first horizontal, then vertical
- "S" for other steps, contrary to "s"
- "n" for no plotting.

Plot symbols (plot characters)

Plot symbols (plot characters)

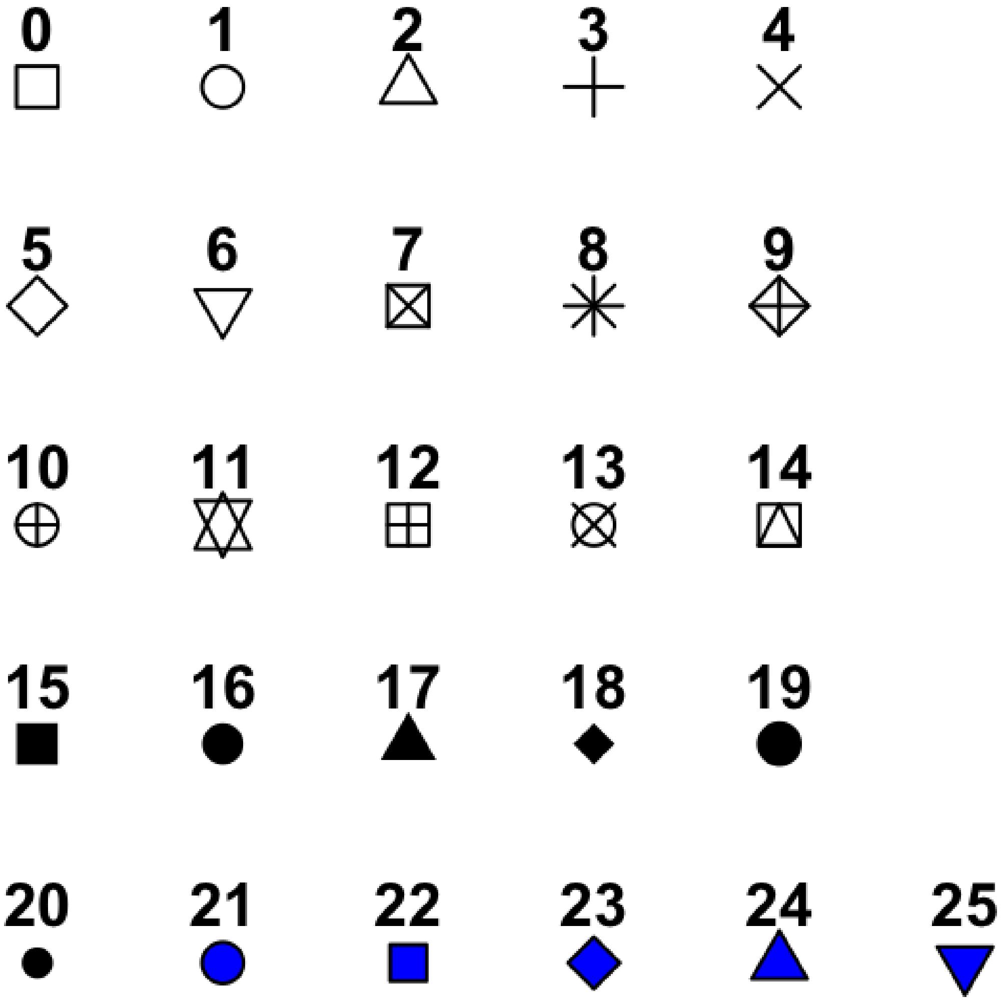
pch = 0,square
pch = 1,circle
pch = 2,triangle point up
pch = 3,plus
pch = 4,cross
pch = 5,diamond
pch = 6,triangle point down
pch = 7,square cross
pch = 8,star
pch = 9,diamond plus
pch = 10,circle plus
pch = 11,triangles up and down
pch = 12,square plus
pch = 13,circle cross

Plot symbols (plot characters)

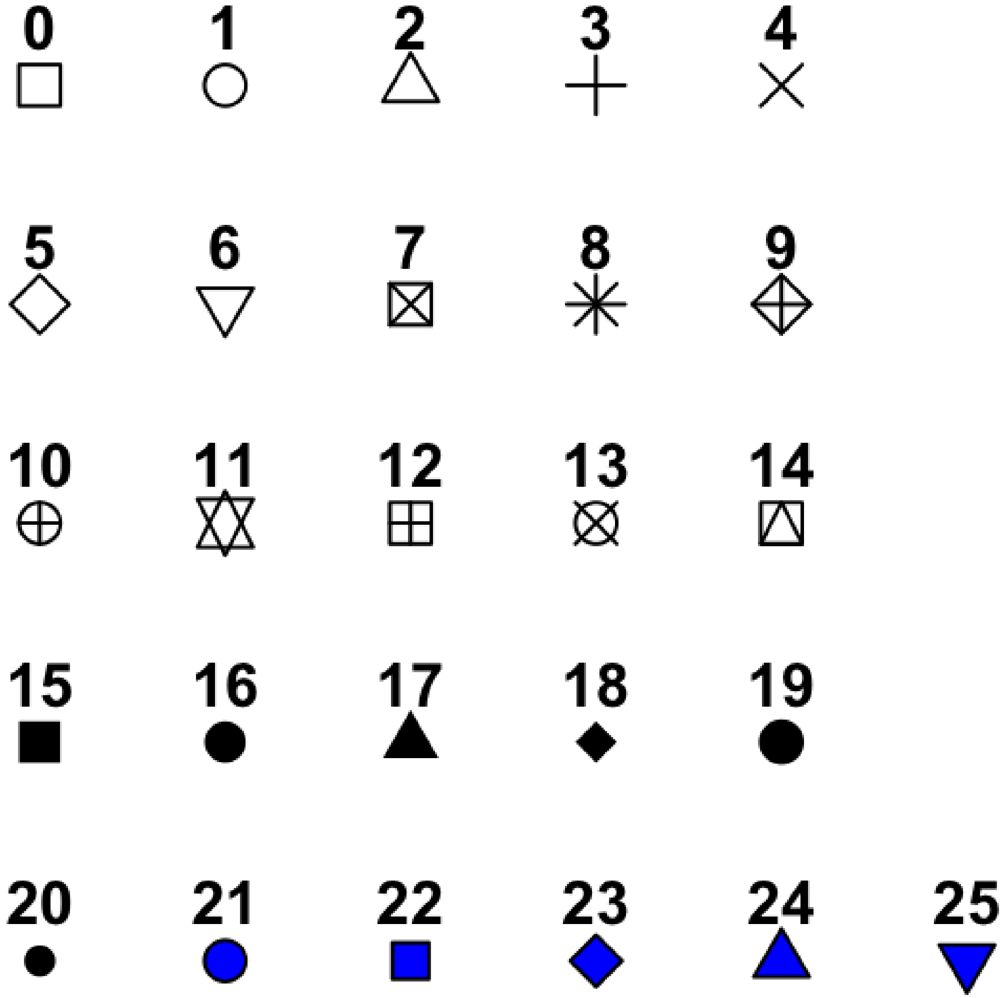
pch = 0,square
pch = 1,circle
pch = 2,triangle point up
pch = 3,plus
pch = 4,cross
pch = 5,diamond
pch = 6,triangle point down
pch = 7,square cross
pch = 8,star
pch = 9,diamond plus
pch = 10,circle plus
pch = 11,triangles up and down
pch = 12,square plus
pch = 13,circle cross
pch = 14,square and triangle down
pch = 15, filled square
pch = 16, filled circle
pch = 17, filled triangle point-up
pch = 18, filled diamond
pch = 19, solid circle
pch = 20,bullet (smaller circle)
pch = 21, filled circle blue
pch = 22, filled square blue
pch = 23, filled diamond blue
pch = 24, filled triangle point-up blue
pch = 25, filled triangle point down blu

Plot symbols (plot characters)

pch = 0,square
pch = 1,circle
pch = 2,triangle point up
pch = 3,plus
pch = 4,cross
pch = 5,diamond
pch = 6,triangle point down
pch = 7,square cross
pch = 8,star
pch = 9,diamond plus
pch = 10,circle plus
pch = 11,triangles up and down
pch = 12,square plus
pch = 13,circle cross
pch = 14,square and triangle down
pch = 15, filled square
pch = 16, filled circle
pch = 17, filled triangle point-up
pch = 18, filled diamond
pch = 19, solid circle
pch = 20,bullet (smaller circle)
pch = 21, filled circle blue
pch = 22, filled square blue
pch = 23, filled diamond blue
pch = 24, filled triangle point-up blue
pch = 25, filled triangle point down blu


0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Plot symbols: PCH



More on symbols (programming)

More on symbols (programming)

& - ampersand

‘ - apostrophe or single quote

* - asterisk

@ - at

{} - braces or curly brackets

[] - brackets

^ - carat

<> - angle brackets or chevron

~ - tilde

| - pipe

- pound

- - hyphen

ggplot()

ggplot()

creates a plot object, layer by layer

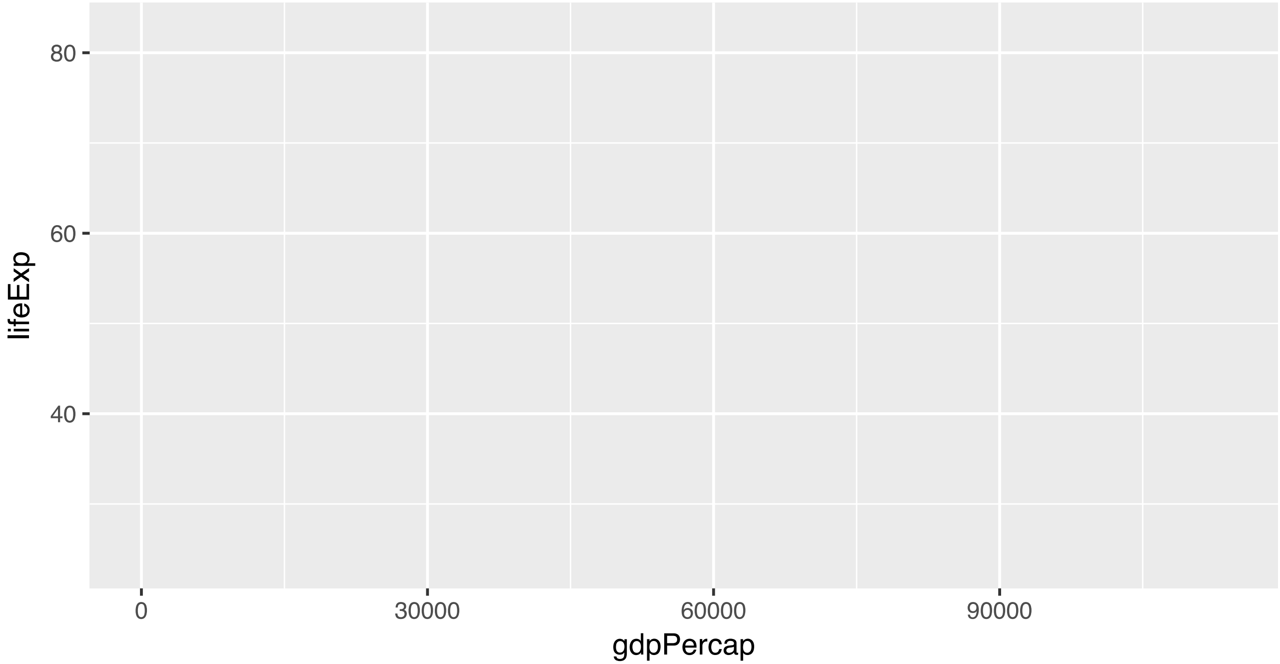
ggplot()

creates a plot object, layer by layer

```
install.packages("ggplot2")
library(ggplot2)
p <- ggplot(data = gm)
p + geom_point(size=2)
p <- ggplot(data = gm,
           mapping = aes(x = gdpPercap,
                         y = lifeExp))
```

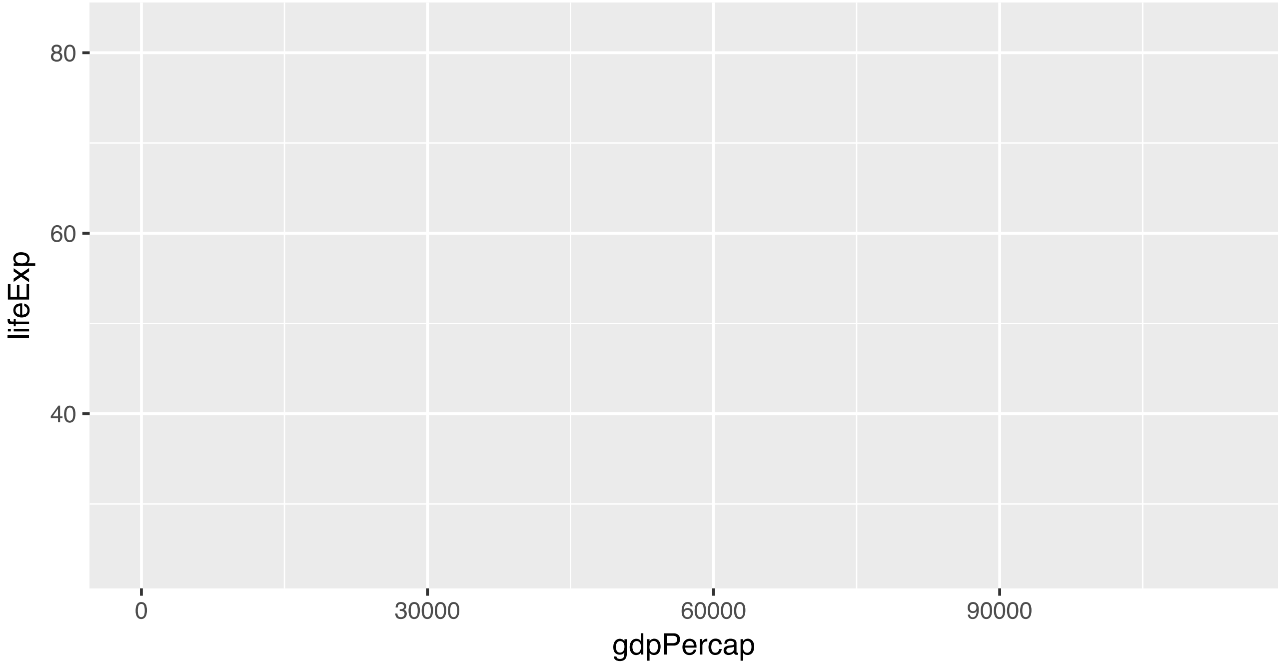
ggplot()

creates a plot object, layer by layer



ggplot()

creates a plot object, layer by layer



plot object p cannot be displayed without adding at least one layer at this point, there is nothing to see!

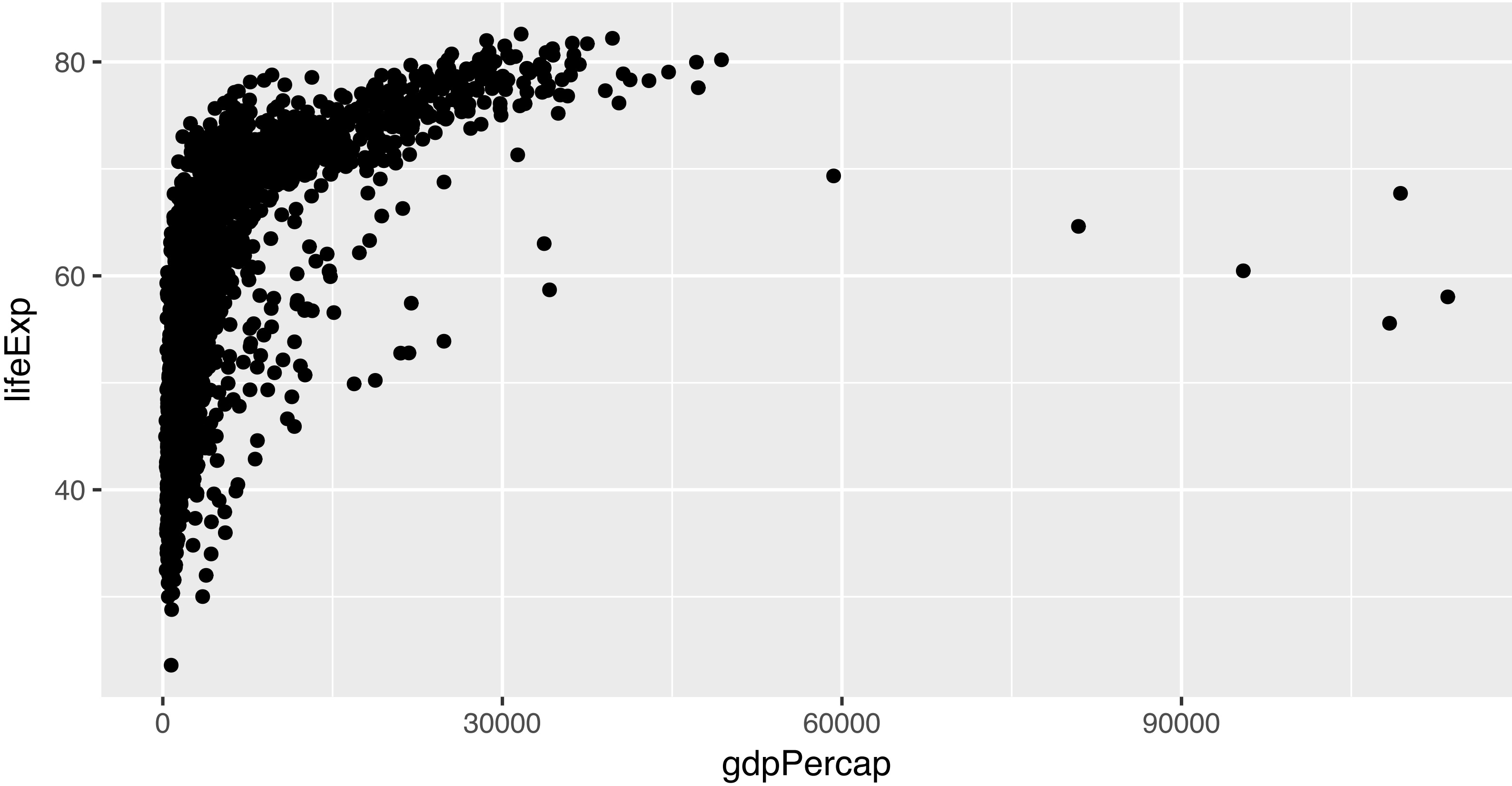
Add a Layer

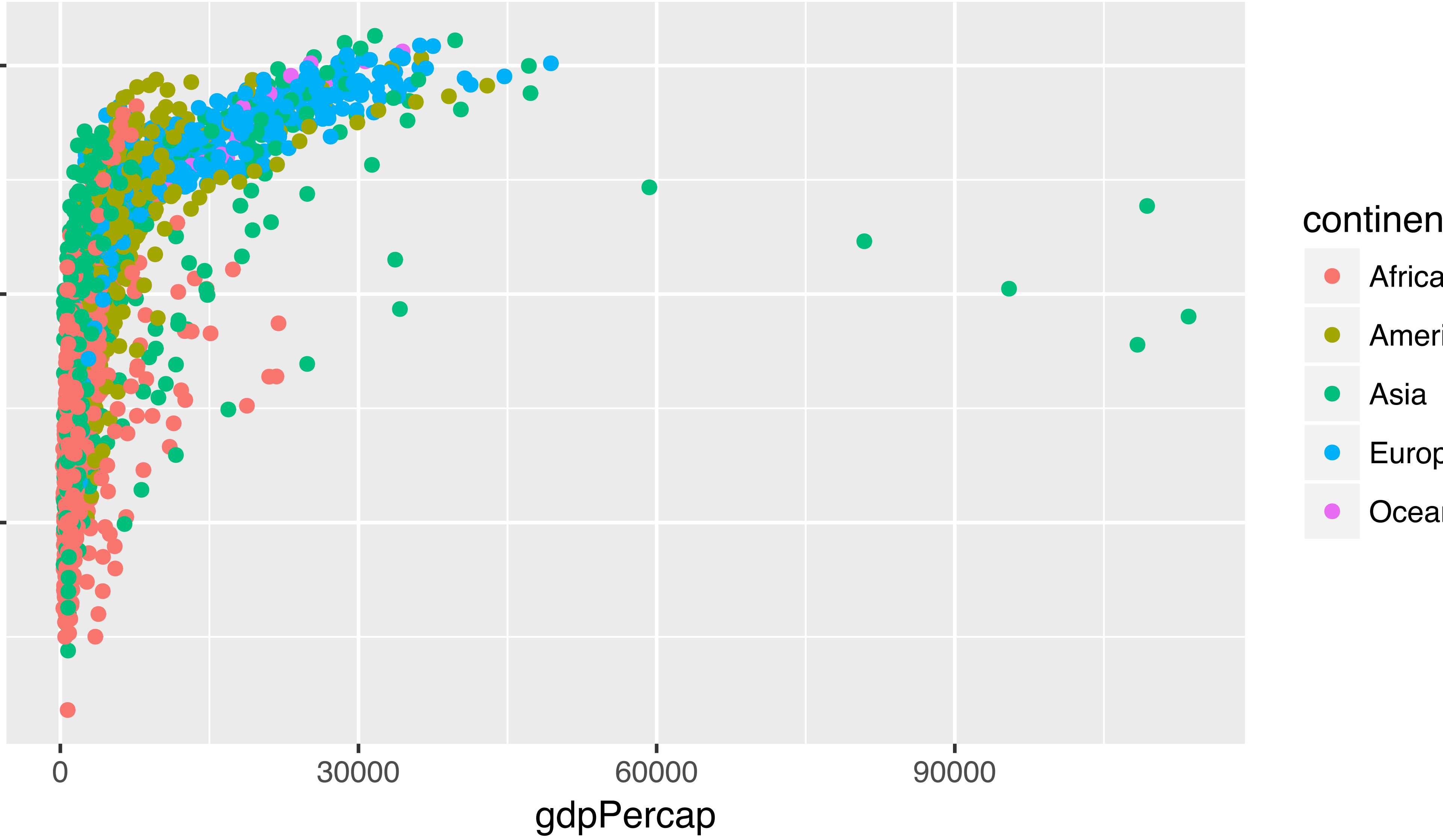
Add a Layer

```
p + geom_point()
```

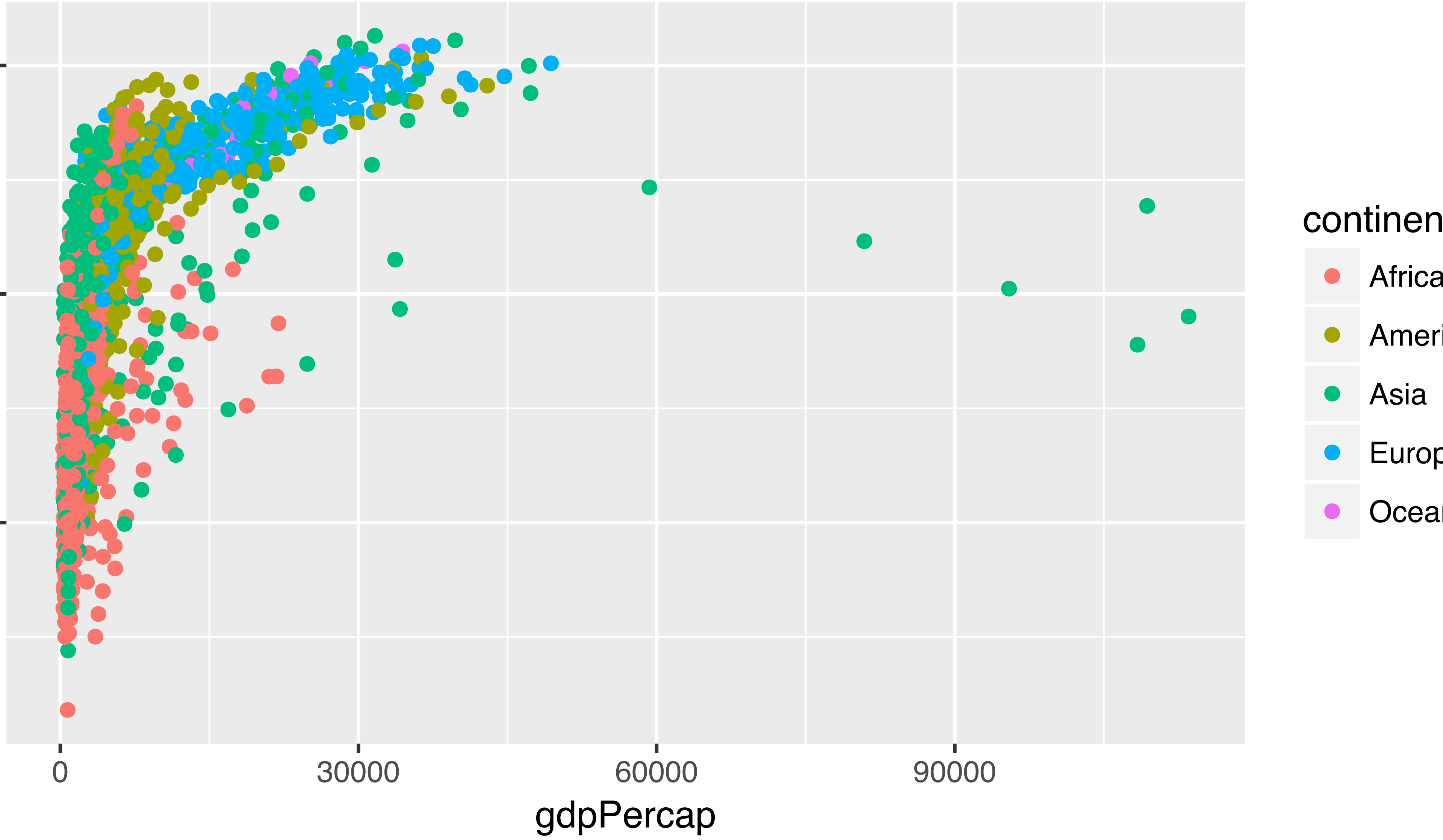
Add a Layer

p + geom_point()

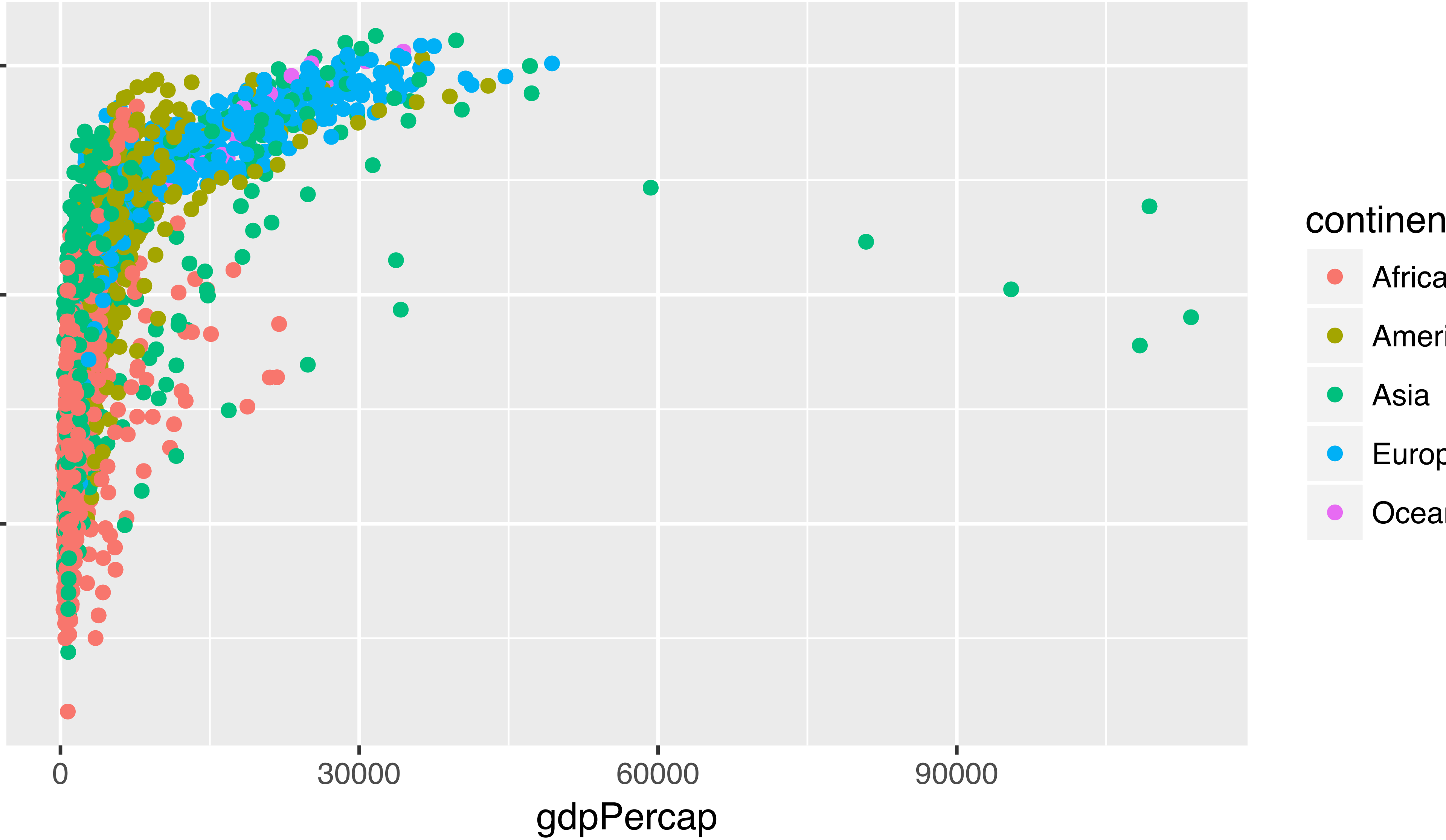


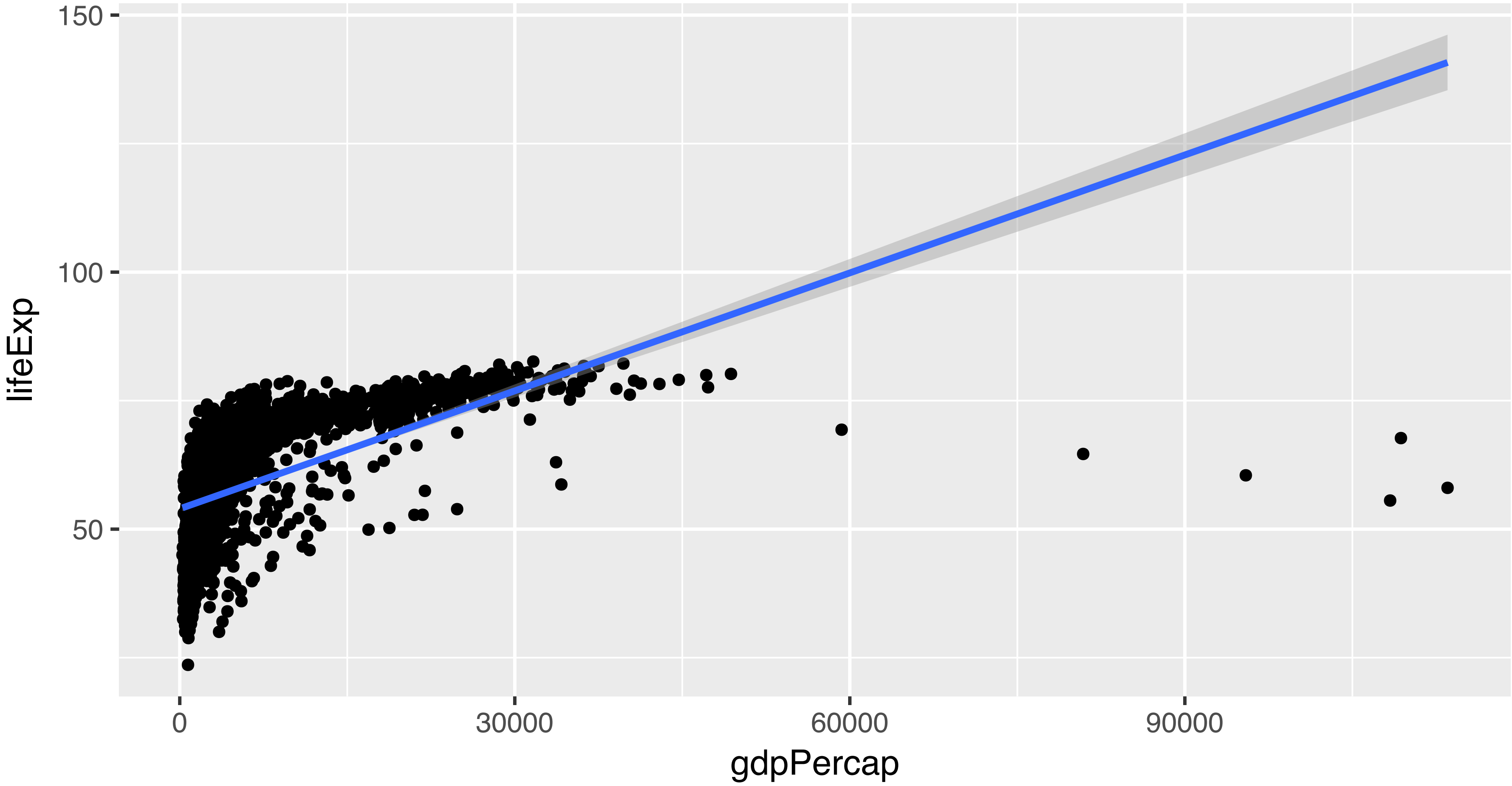


0
30000
60000
90000
gdpPercap
continen
Africa
Ameri
Asia
Europ
Ocear

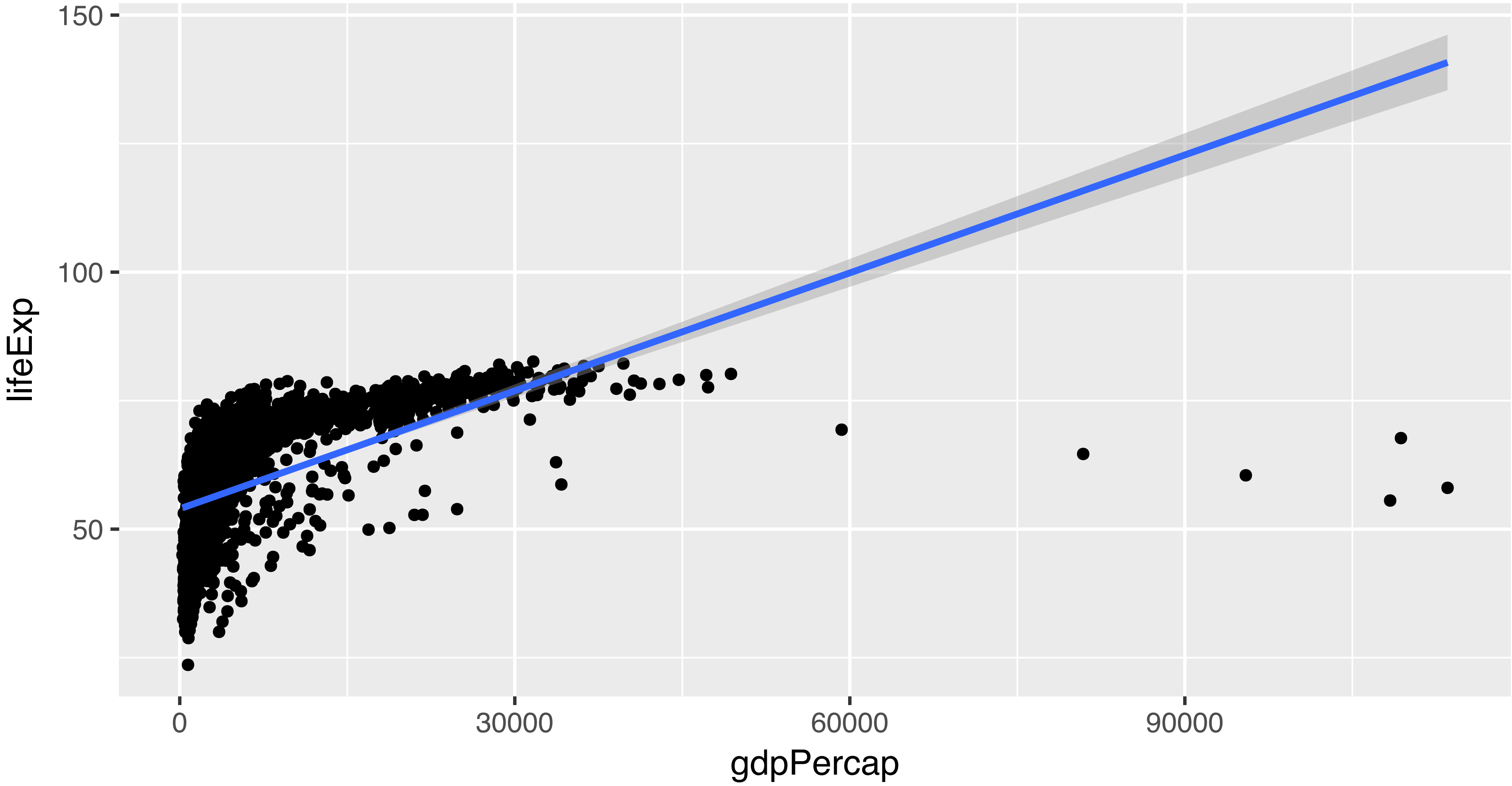


0
30000
60000
90000
gdpPercap
continen
Africa
Ameri
Asia
Europ
Ocean

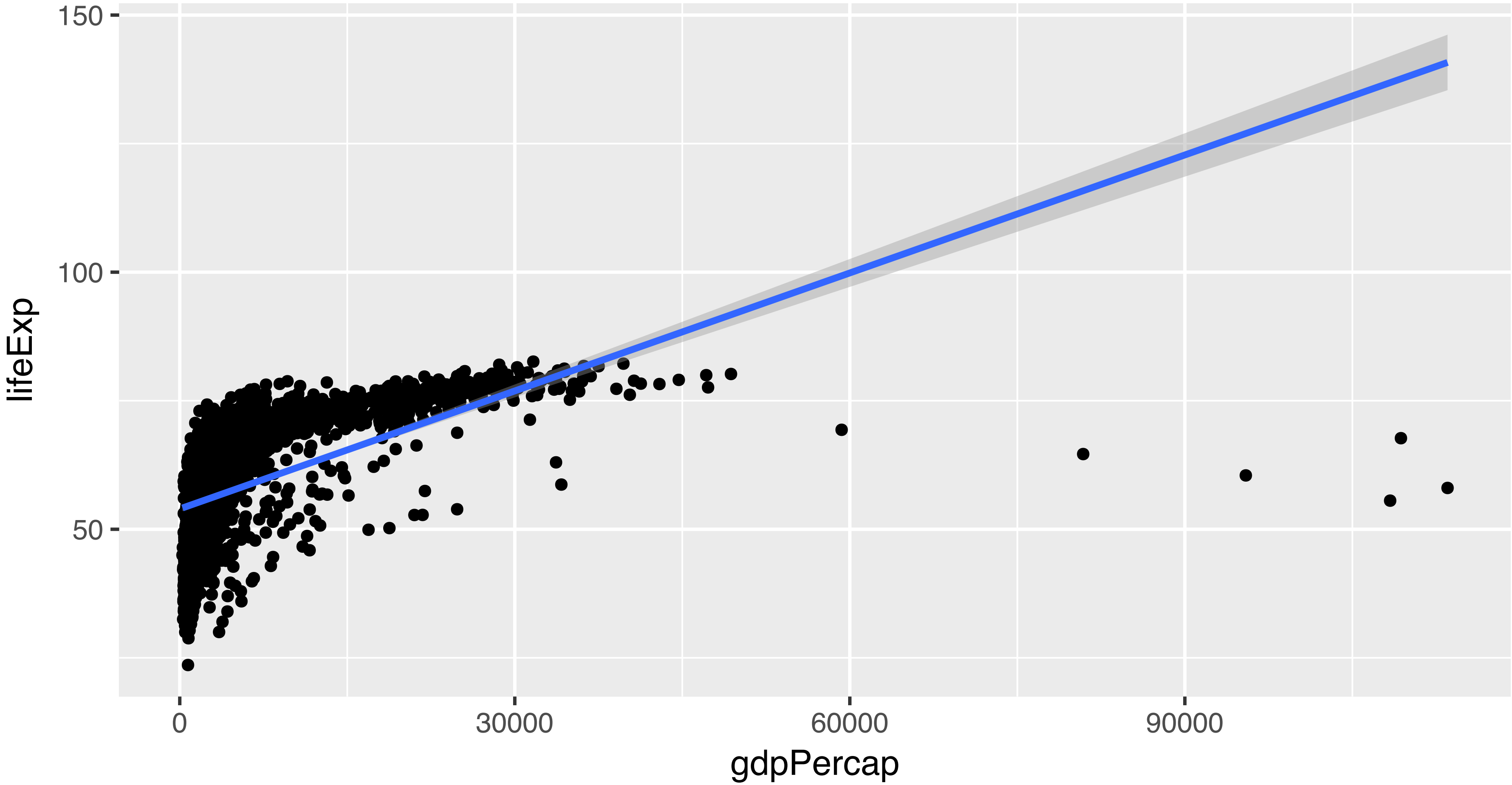

0
30000
60000
90000
gdpPercap
continen
Africa
Ameri
Asia
Europ
Ocean


150
100
50
lifeExp
0
30000
60000
90000
gdpPercap

Add regression line



Add regression line

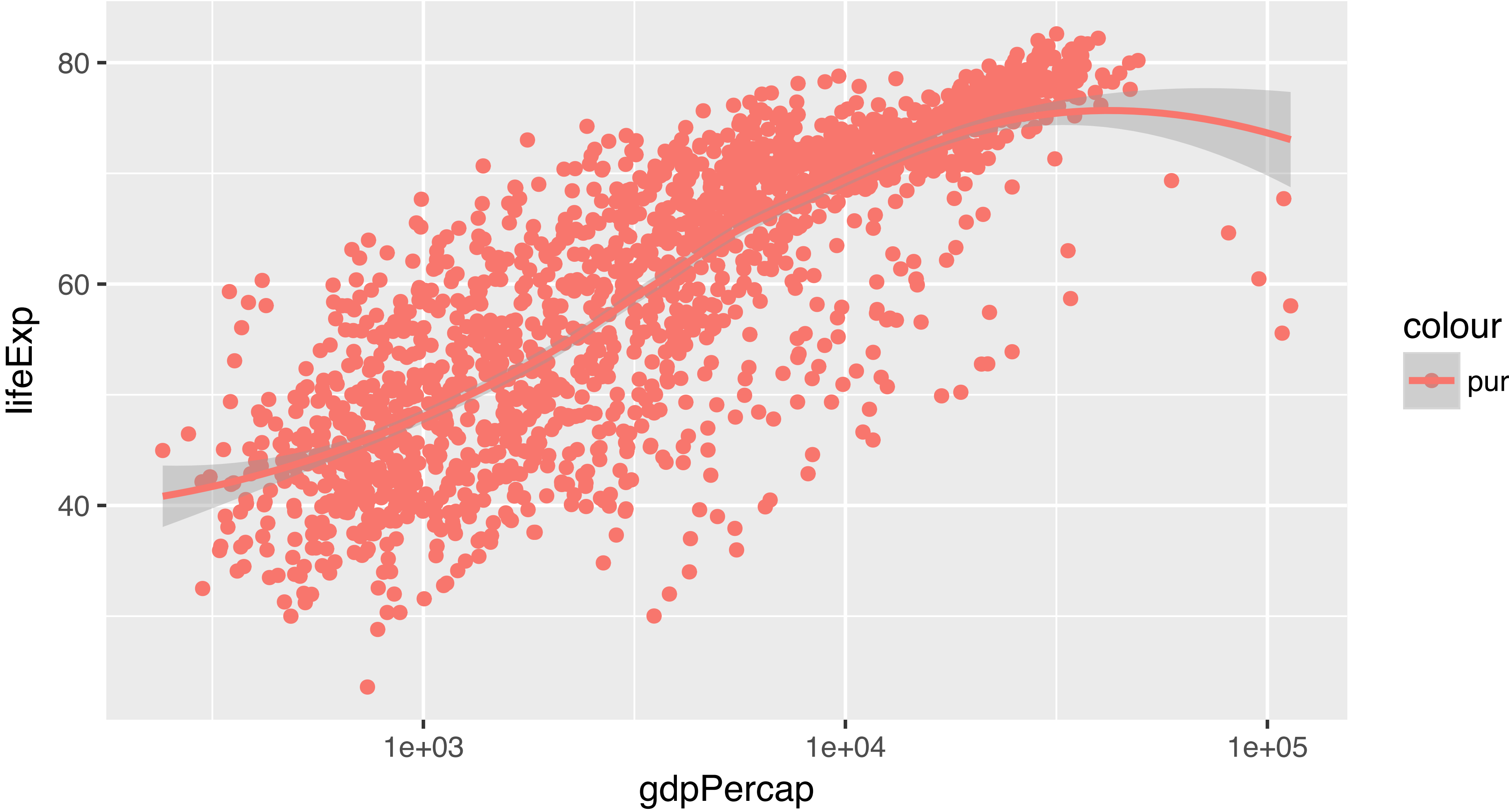


Change dot color?

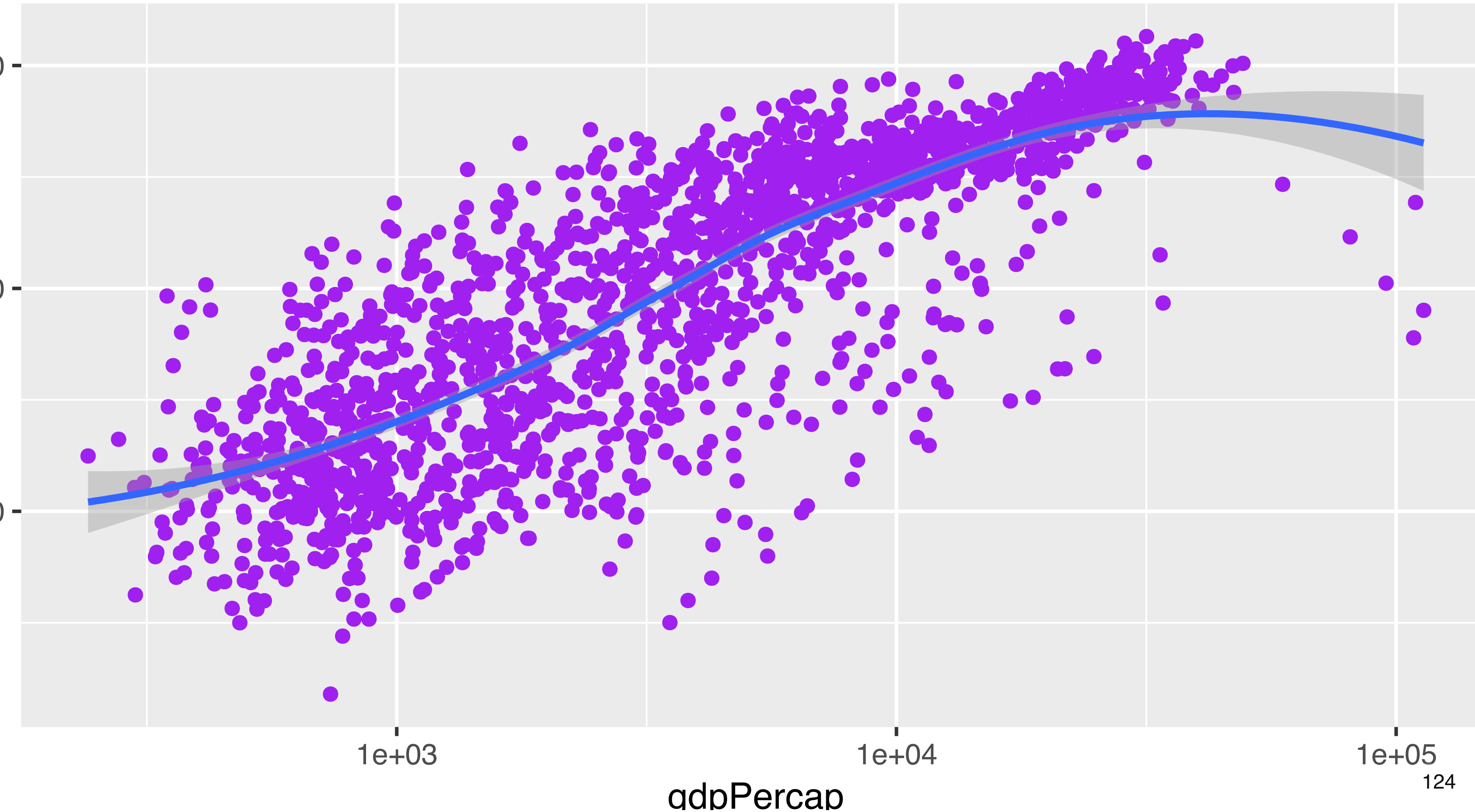
Change dot color?

```
p <- ggplot(data = gm,
            mapping = aes(x = gdpPercap,
                          y = lifeExp,
                          color = "purple"))
p + geom_point() +
  geom_smooth(method = "loess") +
  scale_x_log10()
```

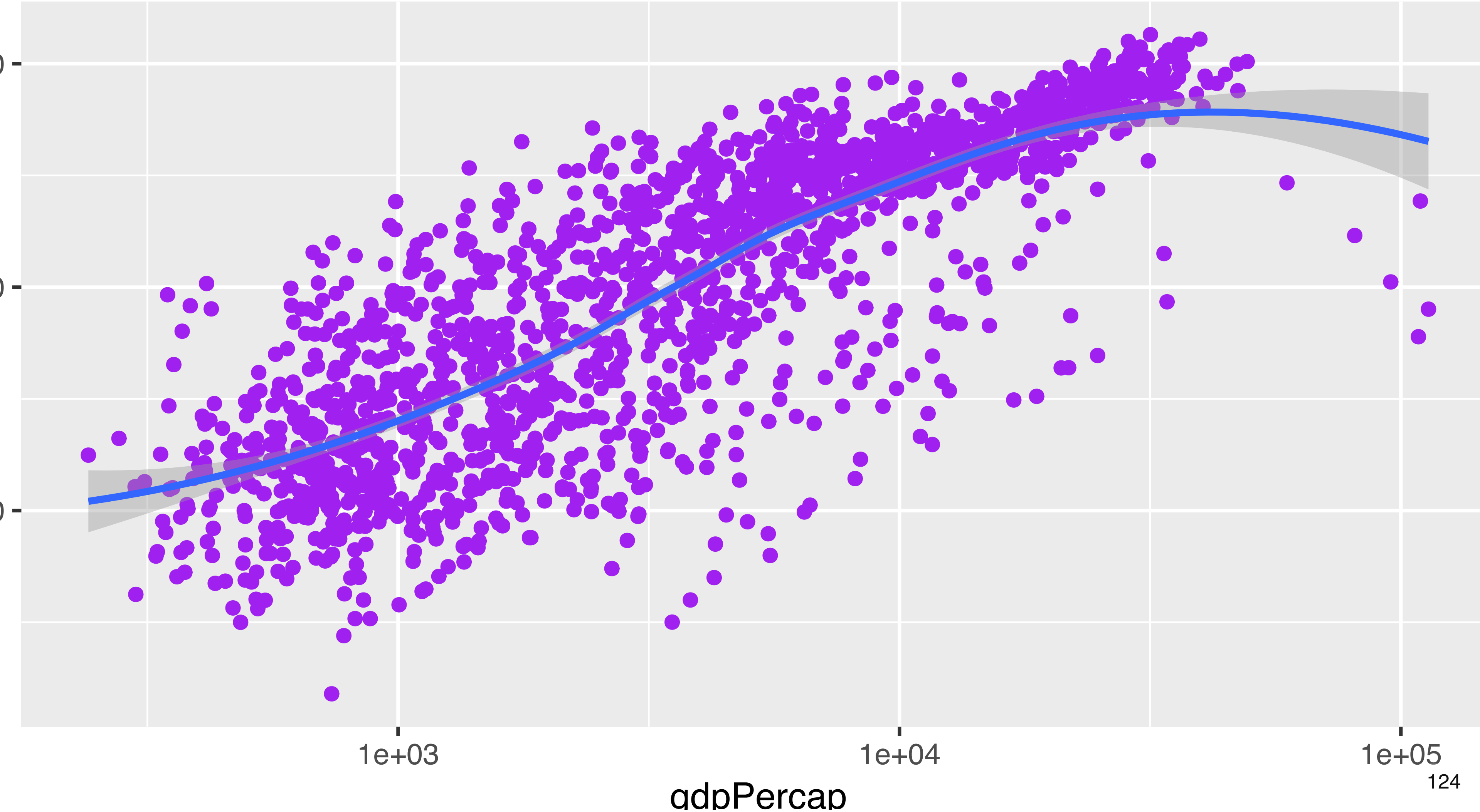
Change dot color?



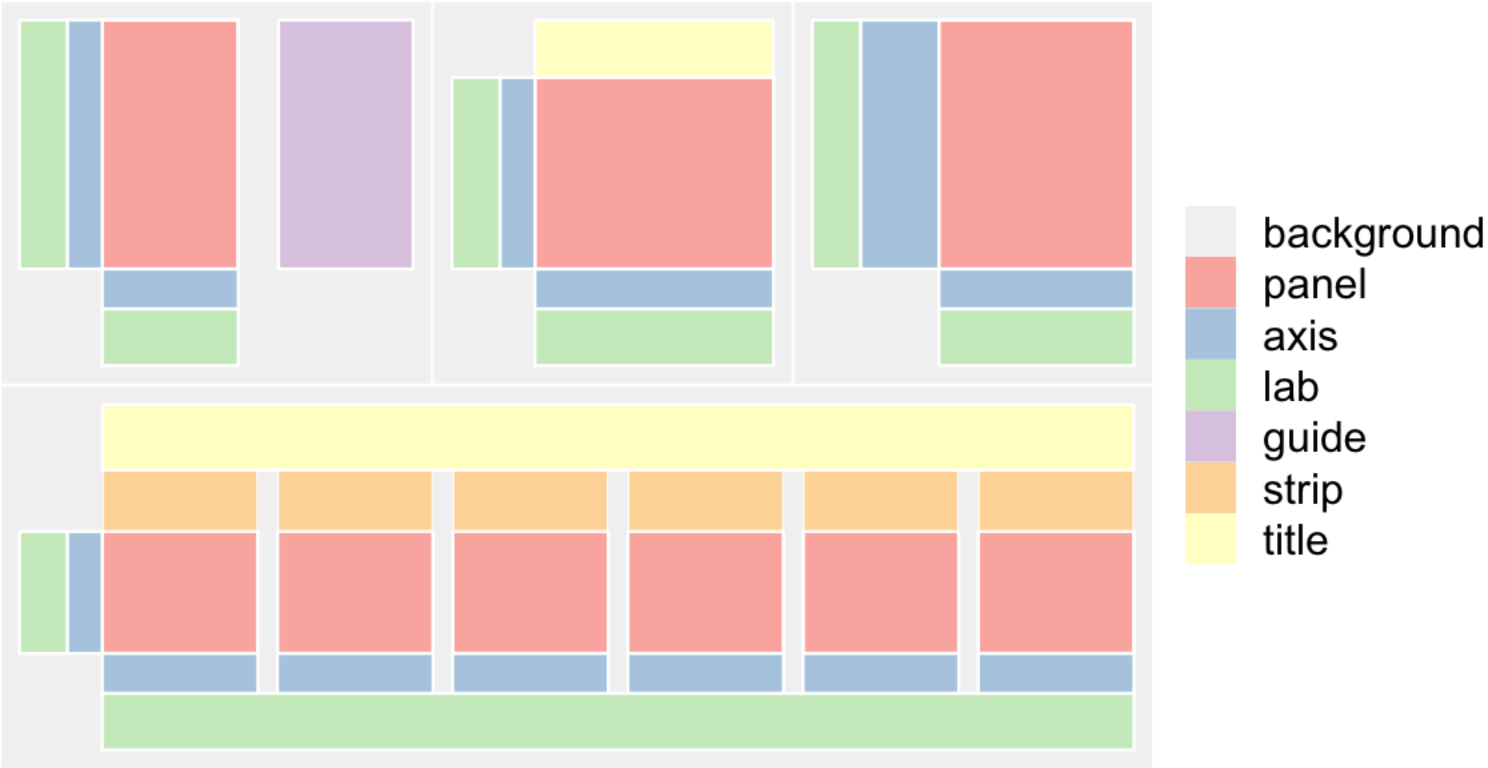
Change dot color?



Change dot color?



ggplot2 multi-plot layout



Source: Baptiste Auguie. 2017. Laying out multiple plots on a page (https://cran.r-project.org/web/packages/egg/vignettes/Ecosystem.html)