Data Visualization

Prof. Dr. Javier Valdes
Javier.valdes@th-deg.de

IAI - Institut für Angewandte Informatik
Institute for Applied Informatics

Technische Hochschule Deggendorf
Technologie Campus Freyung
Grafenauer Str. 22, D- 94078 Freyung, Germany

Tel.: +49 8551 91764 40
Fax: +49 8551 91764 69

Example of Exploratory data Analysis (Peng)

1. Characteristics of exploratory graphs

The goal of making exploratory graphs is usually developing a **personal understanding of the data** and to prioritize tasks for follow up. Details like axis orientation or legends, while present, are generally cleaned up and prettified if the graph is going to be used for communication later. Often color and plot symbol size are used to convey various dimensions of information.

2 Air Pollution in the United States

In this notebook the key question we are interested is: **Are there any counties in the U.S. that exceed the national standard for fine particle pollution?** This question has important consequences because counties that are found to be in violation of the national standards can face serious legal consequences. In particular, states that have counties in violation of the standards are required to create a State Implementation Plan (SIP) that shows how those counties will come within the national standards within a given period of time.

Exploratory analysis: Who, what, and how

Who:

Your audience!

You!

What

Action

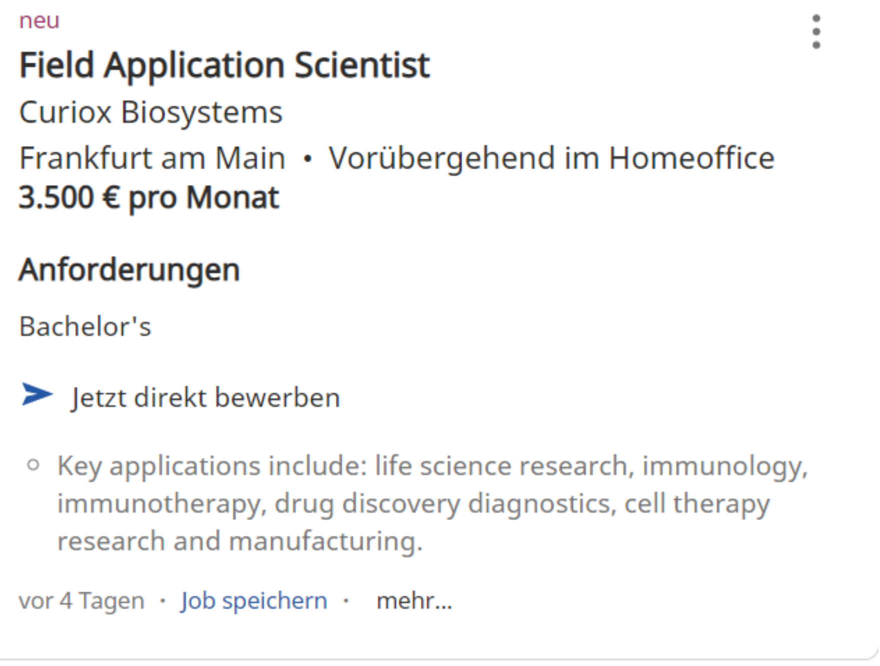
Mechanism

Tone

How

Data

Graphs!



neu

Field Application Scientist

Curiox Biosystems

Frankfurt am Main • Vorübergehend im Homeoffice

3.500 € pro Monat

Anforderungen

Bachelor's

Jetzt direkt bewerben

- Key applications include: life science research, immunology, immunotherapy, drug discovery diagnostics, cell therapy research and manufacturing.

vor 4 Tagen · Job speichern · mehr...

Example: Vaccination

Who: The budget committee that can approve funding for continuation of the winter vaccination program.

What: The vaccination program was a success; please approve budget of $X to continue.

How: illustrate success with data collected through the survey conducted before and after the pilot program.

Exploratory analysis: Who, what, and how

Who:

Your audience!

You!

What

Action

Mechanism

Tone

How

Data

Graphs!

neu

Field Application Scientist

Curiox Biosystems

Frankfurt am Main • Vorübergehend im Homeoffice

3.500 € pro Monat

Anforderungen

Bachelor's

Jetzt direkt bewerben

- Key applications include: life science research, immunology, immunotherapy, drug discovery diagnostics, cell therapy research and manufacturing.

vor 4 Tagen · Job speichern · mehr...

Example: Vaccination

Who: The budget committee that can approve funding for continuation of the winter vaccination program.

What: The vaccination program was a success; please approve budget of $X to continue.

How: illustrate success with data collected through the survey conducted before and after the pilot program.

Exploratory analysis: Who, what, and how

- What background information is relevant or essential?
- Who is the audience or decision maker? What do we know about them?
- What biases does our audience have that might make them supportive of or resistant to our message?
- What data is available that would strengthen our case? Is our audience familiar with this data, or is it new?
- Where are the risks: what factors could weaken our case and do we need to proactively address them?
- What would a successful outcome look like?
- If you only had a limited amount of time or a single sentence to tell your audience what they need to know, what would you say.

3-minute story

Imagine you are a Field Application Scientist. You just wrapped up an experimental pilot that was aimed at reducing exposure to Influenza in a nursing home. You surveyed the old age persons at the onset and end of the program to understand whether and how perceptions toward vaccination changed. You believe the data shows a great success story. You would like to expand the use of your methodology in other nursing homes to gather more data.

Exploratory analysis: Who, what, and how

3-minute story

Who?	What?	How?
The call to action would be different for the different groups	Show before & after survey data to demonstrate success of program	Data collected via survey at the onset and end of the program

Exploratory analysis: Who, what, and how

3-minute story

Issue:
Elderly people have bad attitudes about vaccination

Demonstrate Issue:
Show Elderly population rates over the course of last campaigns

Ideas for overcoming issue, including developing pilot program

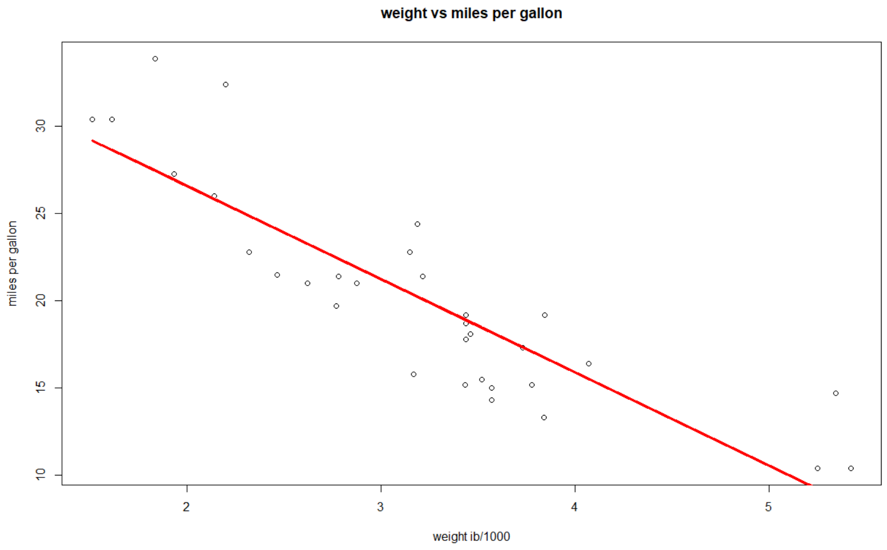
Describe pilot program-goals, structure etc.

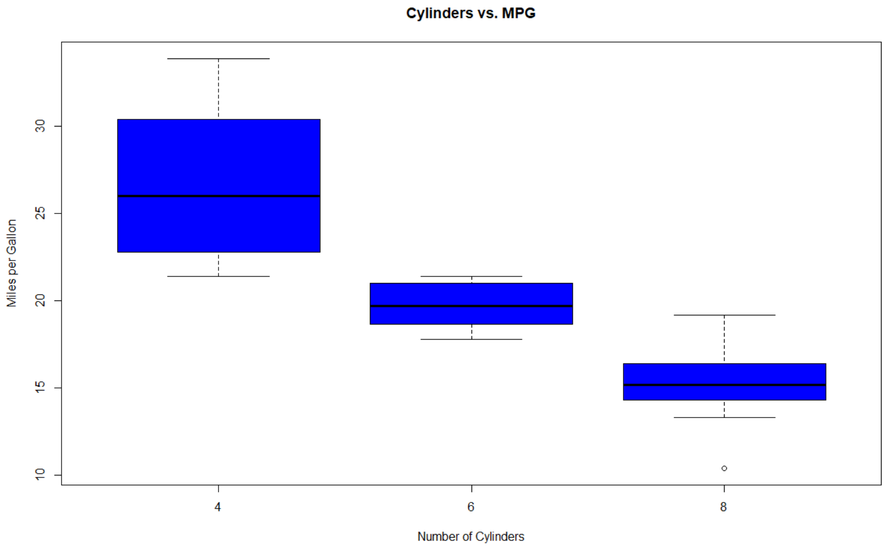
Show before & after survey data to demonstrate success of program

RECOMMENDATION:
Pilot was a success, let's expand it we need $$

Exploratory analysis: Who, what, and how

3-minute story





Why investing on visual information?

How many P's can you find in the text?

A M C D F G O I S H P O F Q H O P I U O I U L F K S D K

F K F J Z F C P T H B M U G I N D I C A T O R S D H D X

B E W Z C O M P O S I T E A E T F R J L M N O J K P B R

L A D X O G F J E I L K S A P R P U E D G H M X O Q B I

Effective communication through visualization

How many P's can you find in the text?

A M C D F G O I S H P O F Q H O P I U O I U L F K S D K

F K F J Z F C P T H B M U G I N D I C A T O R S D H D X

B E W Z C O M P O S I T E A E T F R J L M N O J K P B R

L A D X O G F J E I L K S A P R P U E D G H M X O Q B I

Effective communication through visualization

Article Talk Read Edit View history Search Wikipedia

Principles of grouping

From Wikipedia, the free encyclopedia

This article has multiple issues. Please help improve it or discuss these issues on the **talk page**. *(Learn how and when to remove these template messages)* [hide]

- This article **needs additional citations for verification**. *(December 2018)*
- This article **needs attention from an expert in Cognitive science**. *(April 2011)*

The **principles of grouping** (or **Gestalt laws of grouping**) are a set of principles in psychology, first proposed by Gestalt psychologists to account for the observation that humans naturally perceive objects as organized patterns and objects, a principle known as Prägnanz. Gestalt psychologists argued that these principles exist because the mind has an innate disposition to perceive patterns in the stimulus based on certain rules. These principles are organized into five categories: Proximity, Similarity, Continuity, Closure, and Connectedness.[1][2][3][4]

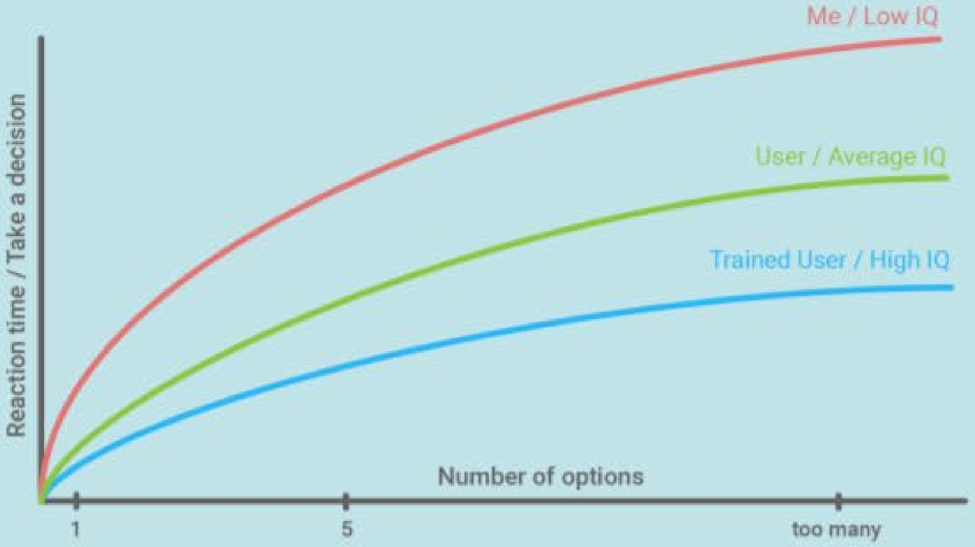
Irvin Rock and Steve Palmer, who are acknowledged as having built upon the work of Max Wertheimer and others and to have identified additional grouping principles,[5] note that Wertheimer's laws have come to be called the "Gestalt laws of grouping" but state that "perhaps a more appropriate description" is "principles of grouping."[6][7] Rock and Palmer helped to further Wertheimer's research to explain human perception of groups of objects and how *whole* objects are formed from *parts* which are perceived.

Contents [hide]

Effective communication through visualization

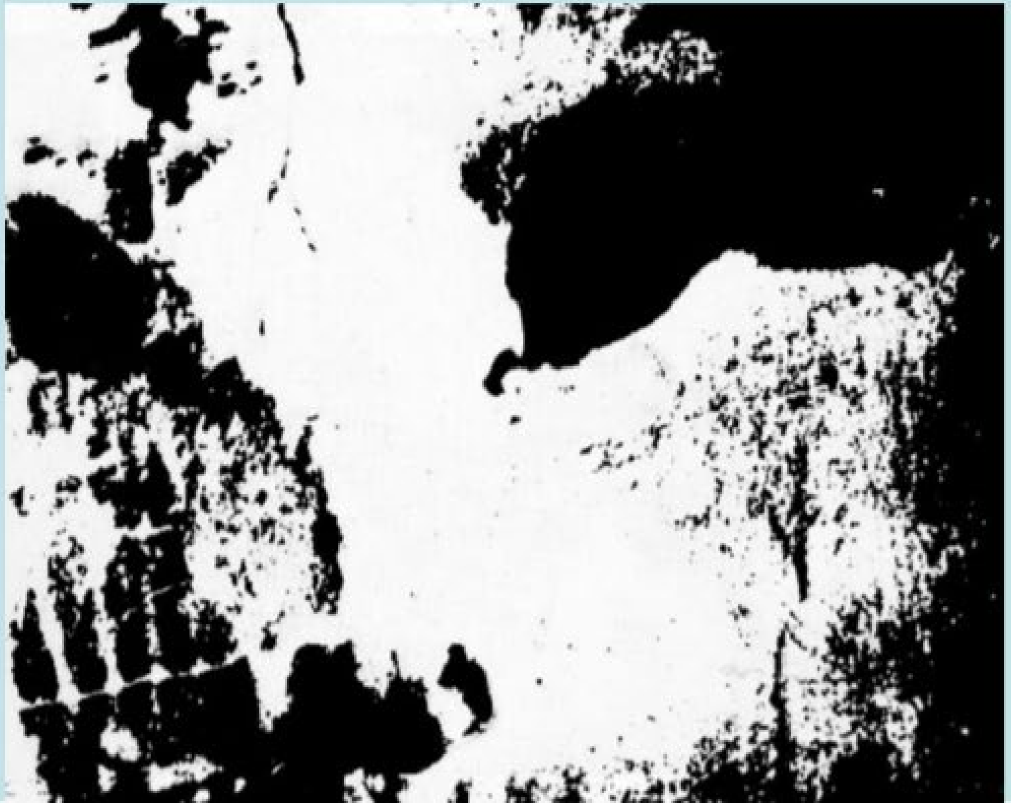
- Gestalt psychology

- Miller's law – The number of objects an average person can hold in working memory is about seven [Also known as the magical number Seven, plus or minus two]
- Hick's law – Describes the time it takes for a person to make a decision as a result of the possible choices avaibale: Increasing the number of choices will increase the decision time logarithmically



Effective communication through visualization

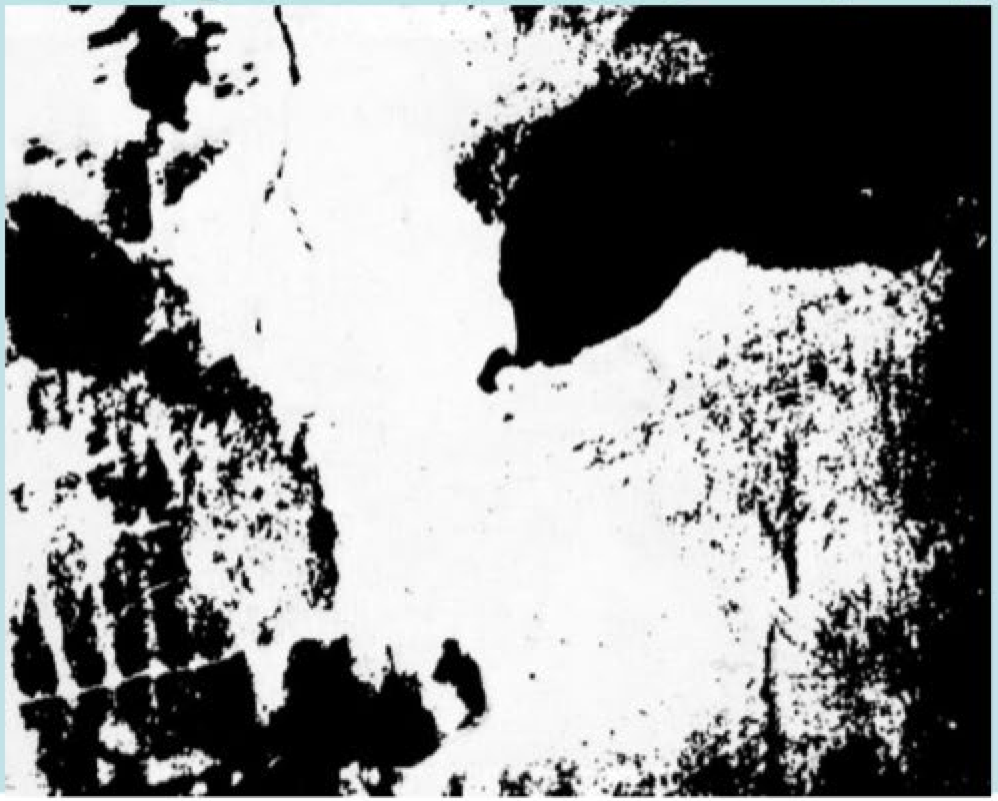
Emergence – forming complex patterns from simple rules.

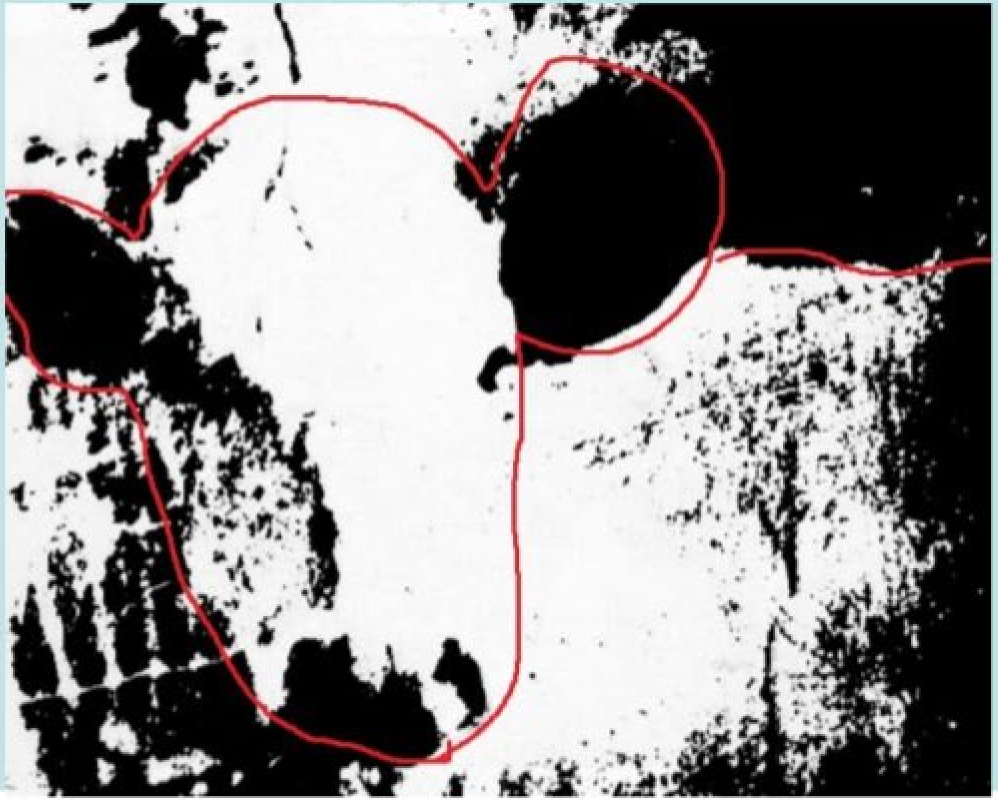


Can you see the cow?

Visual perception and the Principles of Gestalt

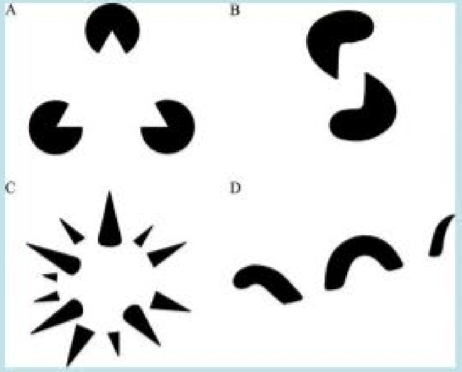
Emergence – forming complex patterns from simple rules.



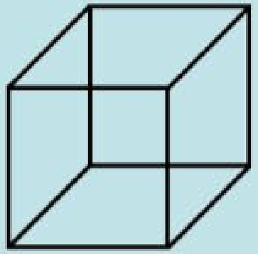


Pragnanz – Simplicity is the key

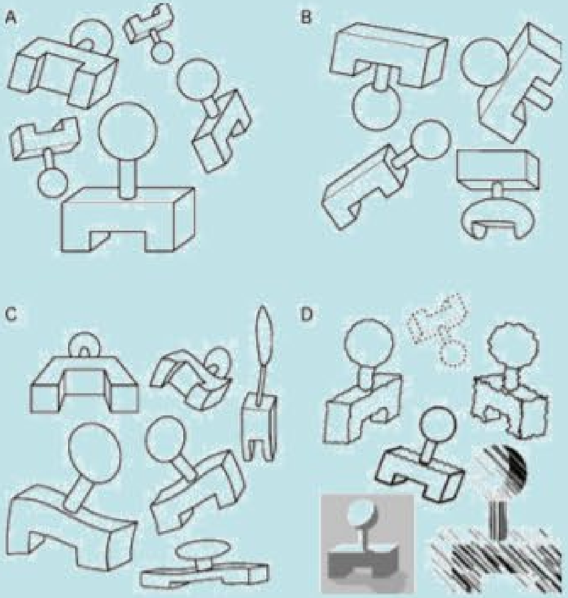
- **Reification -** The perceived experience contains more information than the sensorial stimulus.



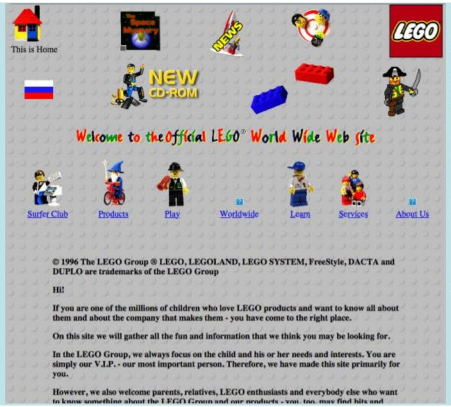
- **Multistability -** Ambiguous perceptual experiences to pop back and forth between alternative interpretations.

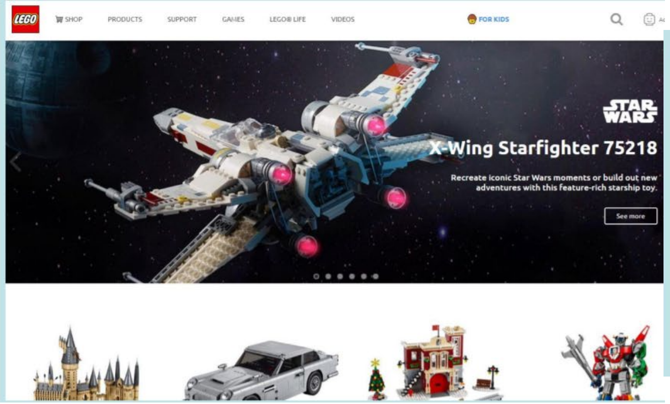


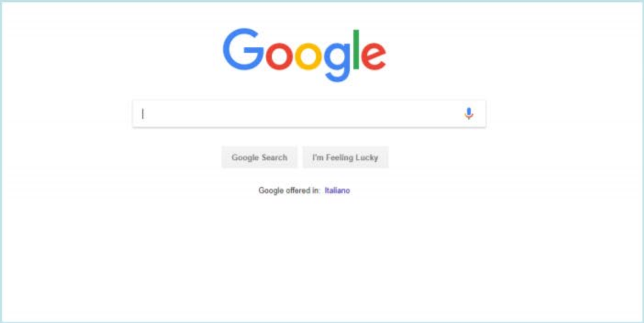
- **Invariance -** Simple objects are recognized independent of rotation, translation and scale.



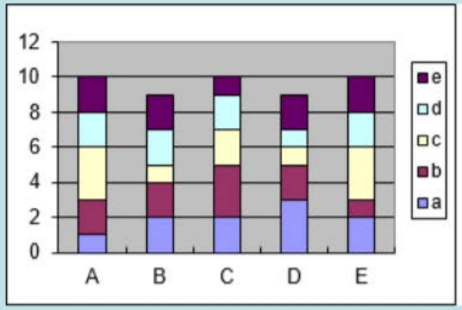
Pragnanz – Simplicity is the key

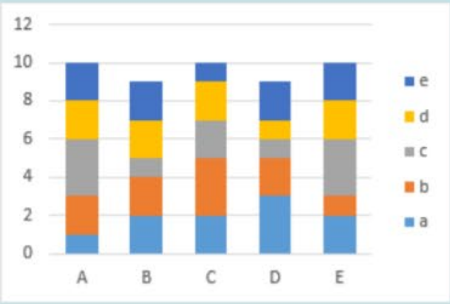






Excel 2003 vs Excel 2013





$$\textit{Data-Ink ratio} = \frac{\textit{data-ink}}{\textit{total ink used to make the graphic}}$$

= 1 – proportion of a graphic that can be erased without loss of data-information

(Tufte, Edward – *The visual display of quantitative information*)

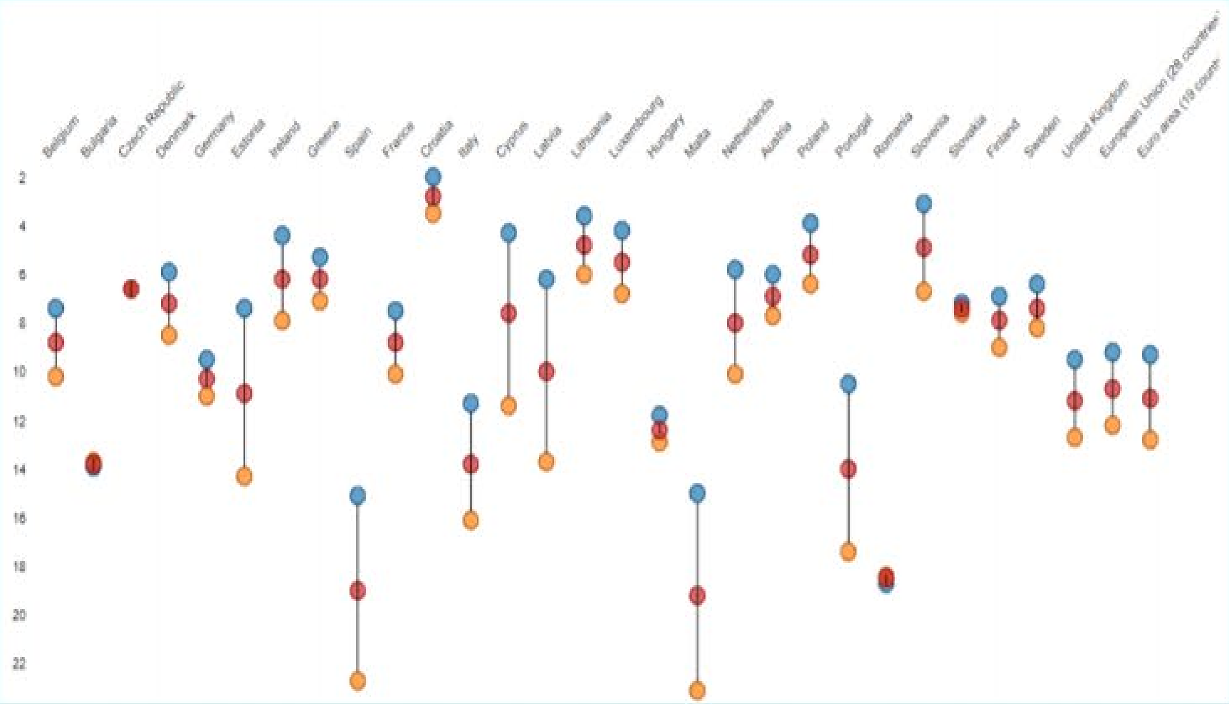
Some suggestions to reduce data-ink ratio:

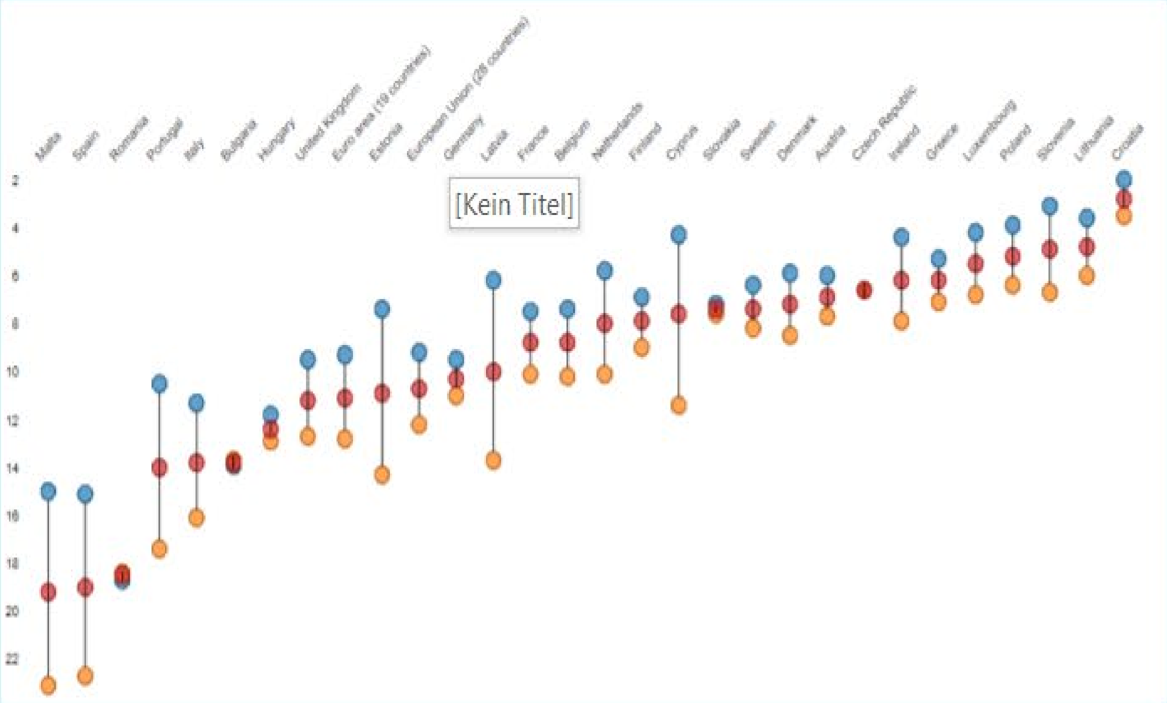
- *No 3d charts*
- *No backgrounds, shadows or gradients*
- *Remove gridlines, decoration, borders, fillcolors*

Effective communication through visualization

What can be removed from a chart while keeping the information?

Effective communication through visualization





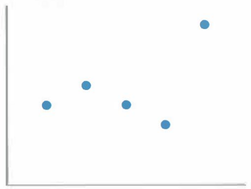
Gestalt principles

- **Law of Simplicity** — Our eyes seek simplicity in complex shapes, preventing us from being overwhelmed by information overload.
- **Law of Proximity** — We perceive objects close to each other as belonging to a group.
- **Law of Similarity** — We seek similarities and differences and link similar items into a group.
- **Law of Figure and Ground** — We tend to segment our visual world into figure and ground. The figure is the object in the focus of our vision, and the ground is the background.
- **Law of Focal Point** — Whatever stands out visually will be given higher attention.

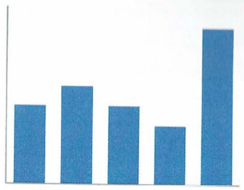
https://medium.com/nightingale/how-to-apply-gestalt-psychology-principles-in-data-visualization-6242f4f1a3de

Choosing an effective visual

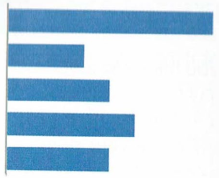
Simple text



Scatterplot



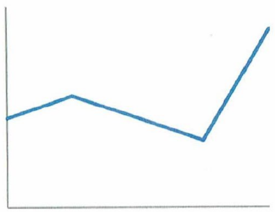
Vertical bar



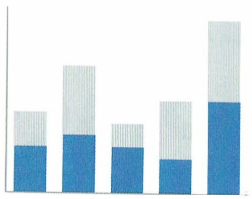
Horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

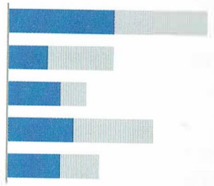
Table



Line



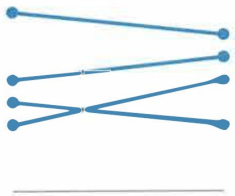
Stacked vertical bar



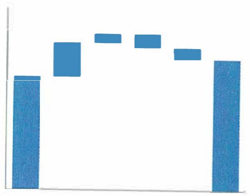
Stacked horizontal bar

	A	B	C
Category 1	15%	22%	
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

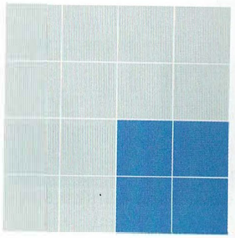
Heatmap



Slopegraph

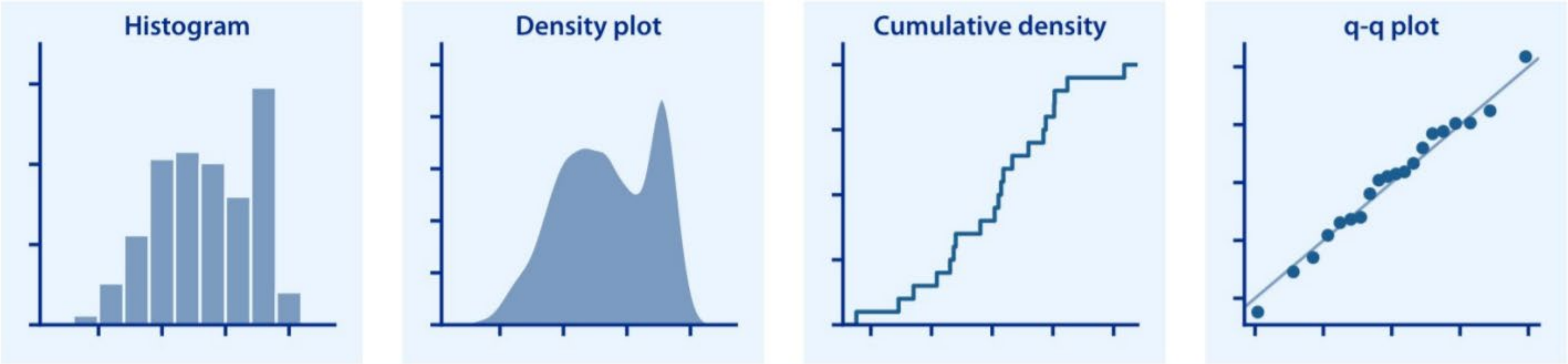


Waterfall

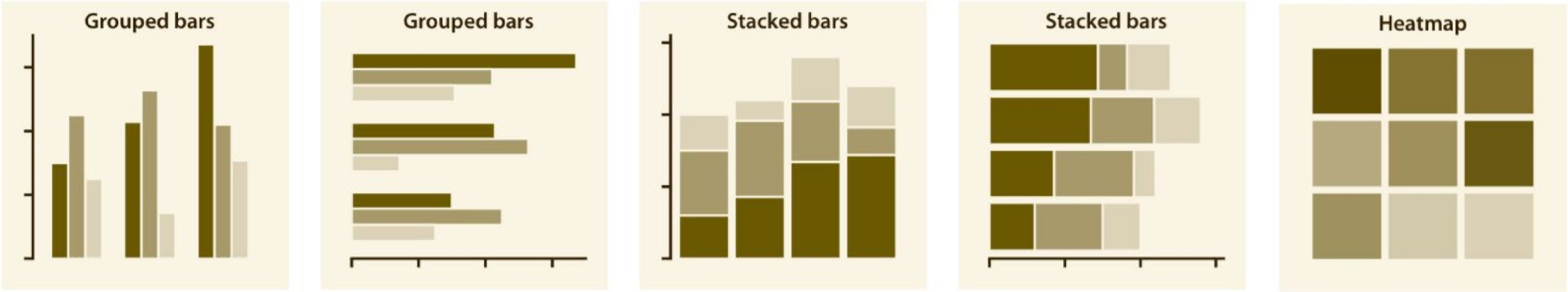


Square area

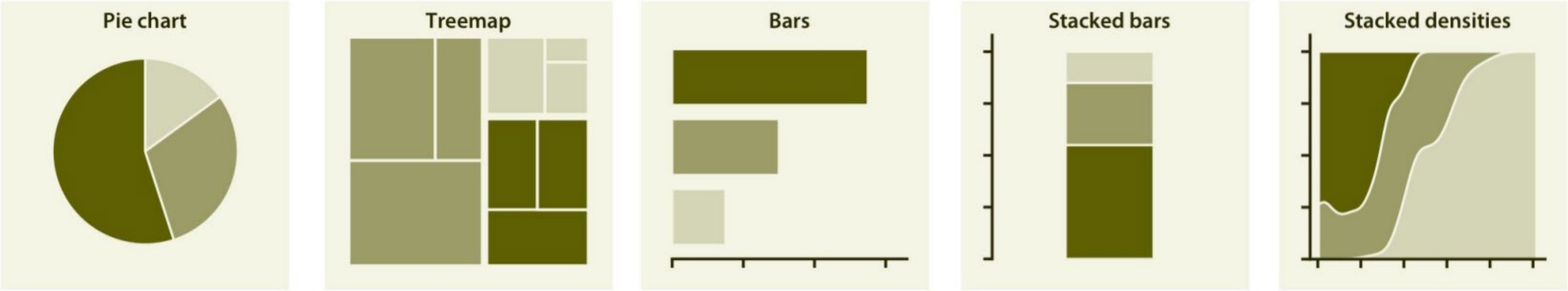
Distributions



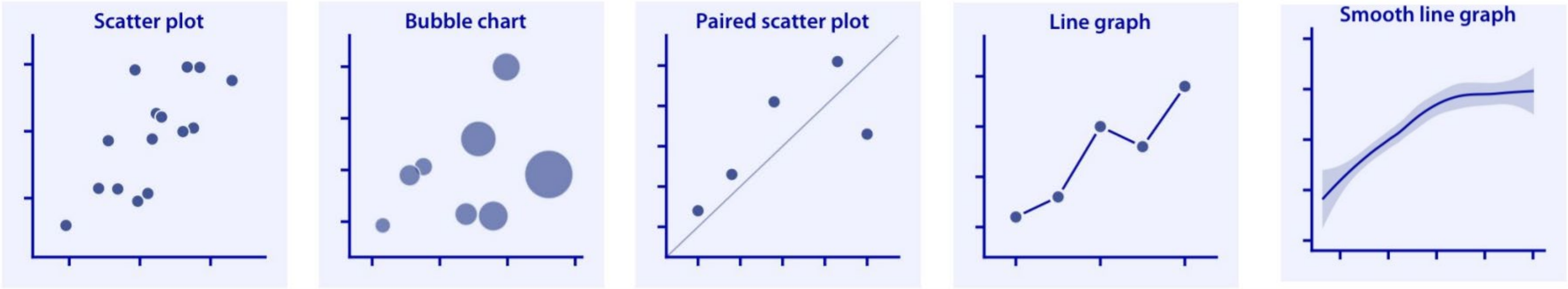
Comparisons



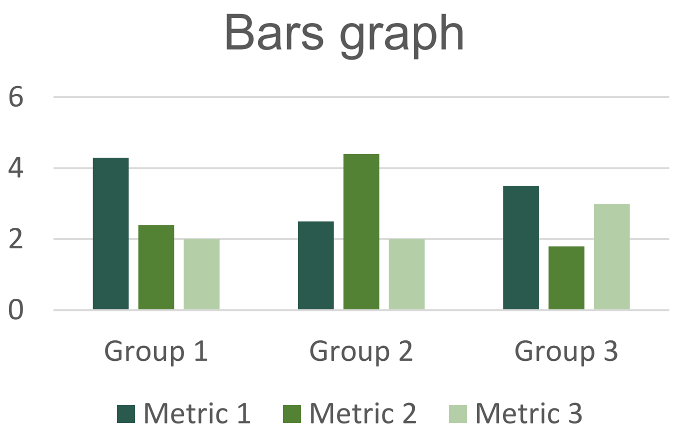
Proportions

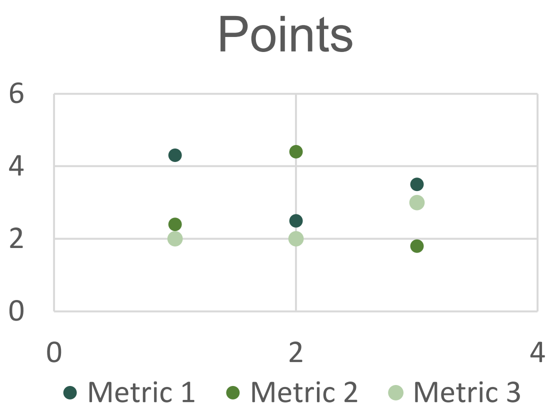


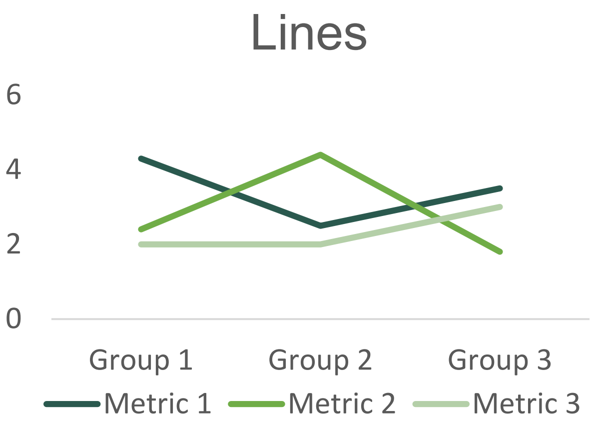
x-y relationships



Graphs







1

While tables interact with our verbal system, **graphs interact with our visual system**, which is **faster** at **processing** information.

2

The types of graphs frequently use fall into **four categories**: points, lines, bars, and area.

3

Note:
there are a plethora of graph types out there, we will cover the **most common**.

Simple text:

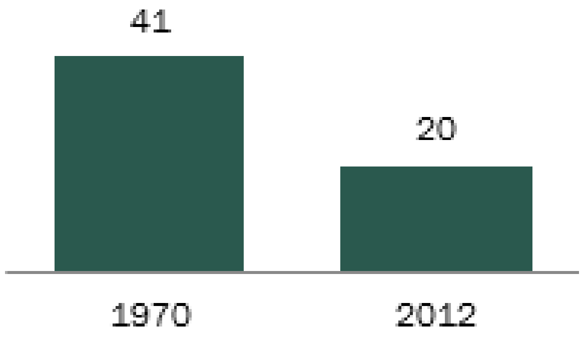
20%

Of children had a
traditional stay-at-home mom
in 2012, compared to 41% in 1970

FIGURE 2.1 Stay-at-home text simple text makeover

Children with a "Traditional" Stay-at-Home Mother

% of children with a married stay-at-home mother with a working husband



Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013

PEW RESEARCH CENTER

Tables

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z.ZZZ
Group 2	$X.X	Y%	Z.ZZZ
Group 3	$X.X	Y%	Z.ZZZ
Group 4	$X.X	Y%	Z.ZZZ
Group 5	$X.X	Y%	Z.ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z.ZZZ
Group 2	$X.X	Y%	Z.ZZZ
Group 3	$X.X	Y%	Z.ZZZ
Group 4	$X.X	Y%	Z.ZZZ
Group 5	$X.X	Y%	Z.ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z.ZZZ
Group 2	$X.X	Y%	Z.ZZZ
Group 3	$X.X	Y%	Z.ZZZ
Group 4	$X.X	Y%	Z.ZZZ
Group 5	$X.X	Y%	Z.ZZZ

1

Tables are great to communicating to a **mixed audience** whose members will each look for their **particular row** of interest

2

The **data** should be what stands out, **not the borders.**

3

Note: Using a table in a live presentation is **rarely** a good idea.

Tables

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z.ZZZ
Group 2	$X.X	Y%	Z.ZZZ
Group 3	$X.X	Y%	Z.ZZZ
Group 4	$X.X	Y%	Z.ZZZ
Group 5	$X.X	Y%	Z.ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z.ZZZ
Group 2	$X.X	Y%	Z.ZZZ
Group 3	$X.X	Y%	Z.ZZZ
Group 4	$X.X	Y%	Z.ZZZ
Group 5	$X.X	Y%	Z.ZZZ

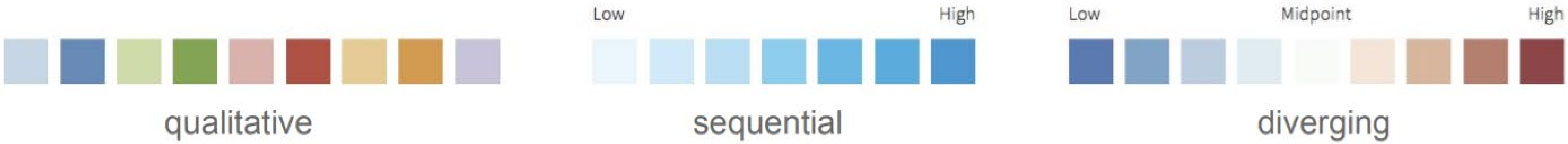
Minimal borders

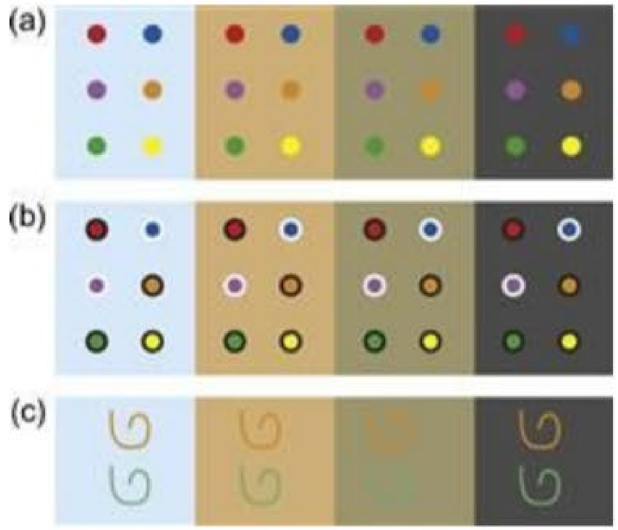
Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z.ZZZ
Group 2	$X.X	Y%	Z.ZZZ
Group 3	$X.X	Y%	Z.ZZZ
Group 4	$X.X	Y%	Z.ZZZ
Group 5	$X.X	Y%	Z.ZZZ

Tables work best when...

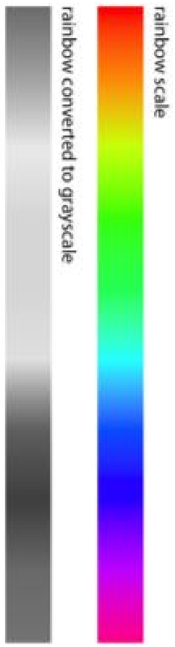
- Used **to look** up individual values
- Used to **compare** individual values
- Data must be **precise**
- You must include **multiple** units of **measure**
- You wish to show both **details** and their **sums**

Effective communication through visualization

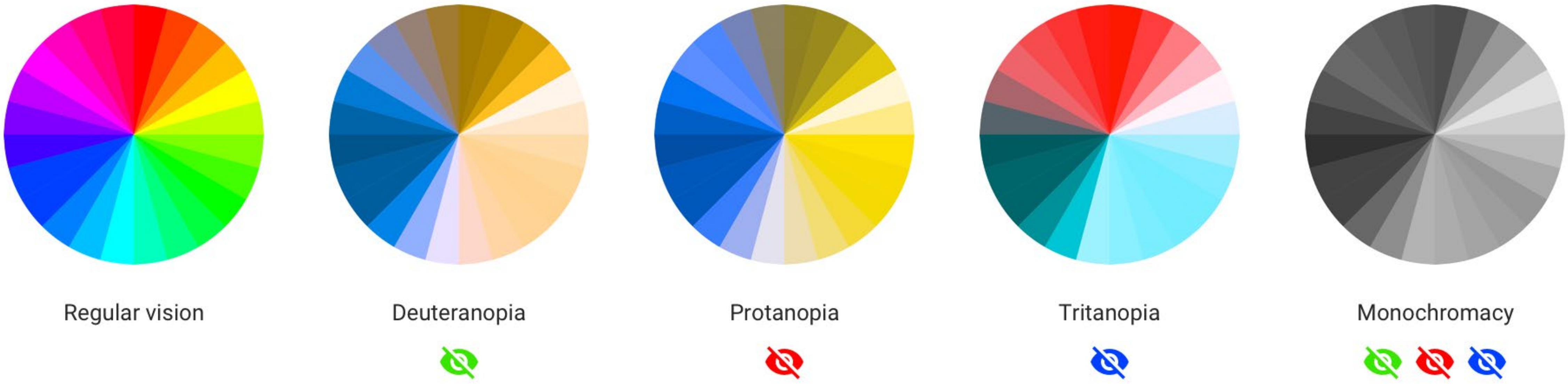




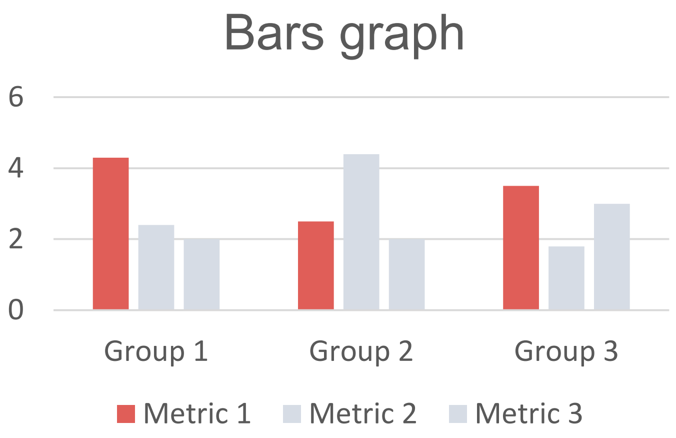
- You need different palettes for continuous, divergent, and discrete data.
- 10 colors **max** for reliable differentiation
- Vary your palette in hue **and** brightness.

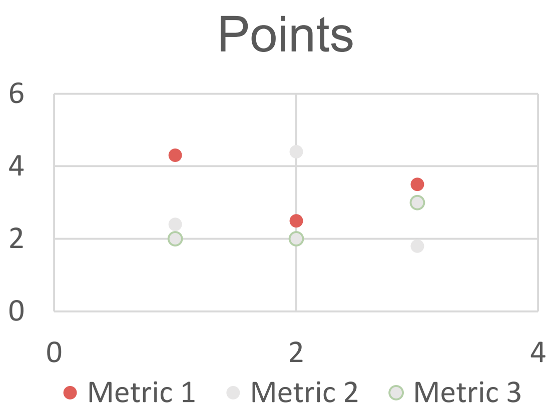


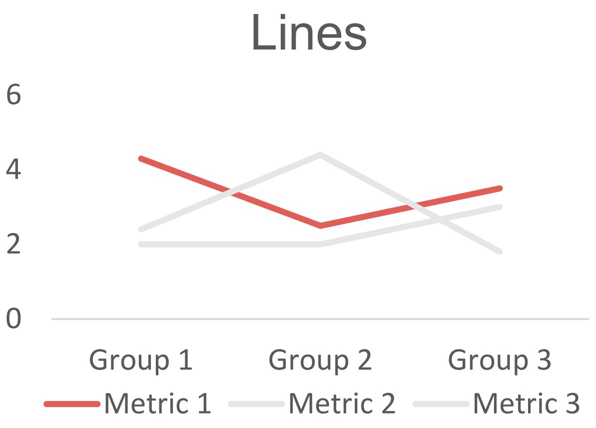
Effective communication through visualization

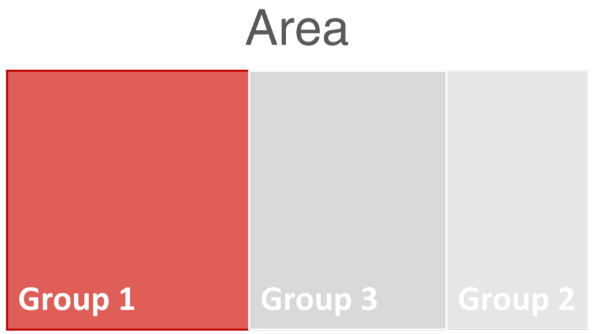


Graphs









1

While tables interact with our verbal system, **graphs interact with our visual system**, which is **faster** at **processing** information.

2

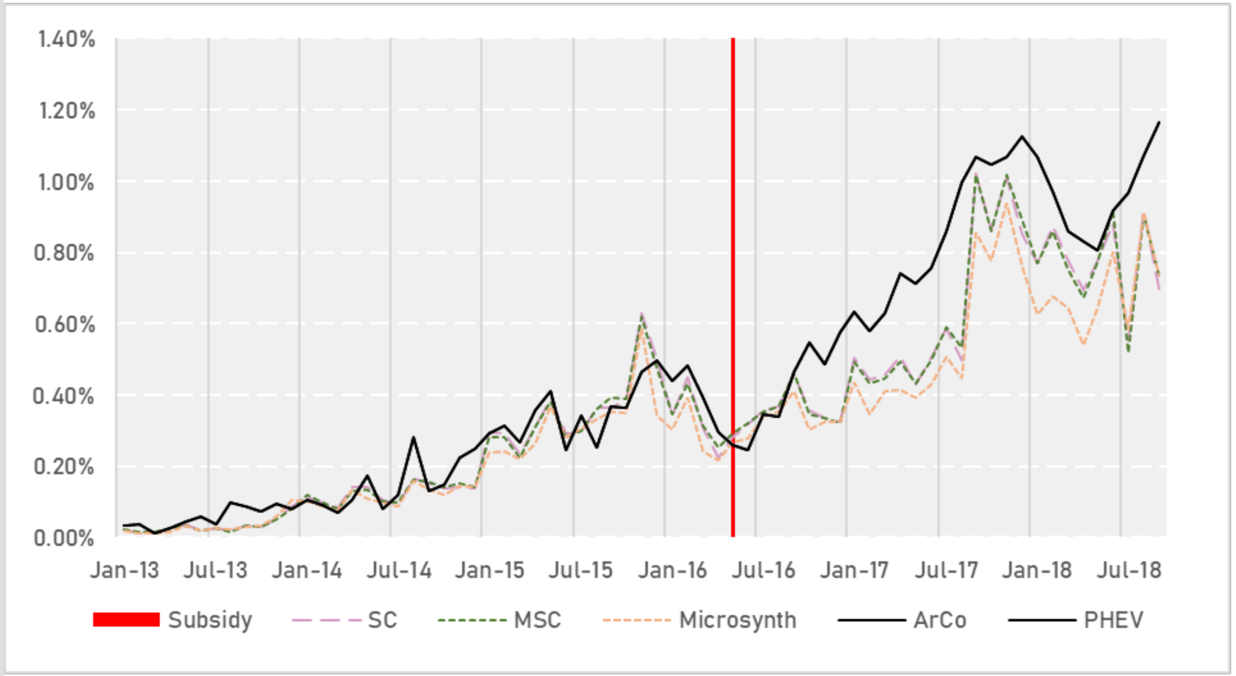
The types of graphs frequently use fall into **four categories**: points, lines, bars, and area.

3

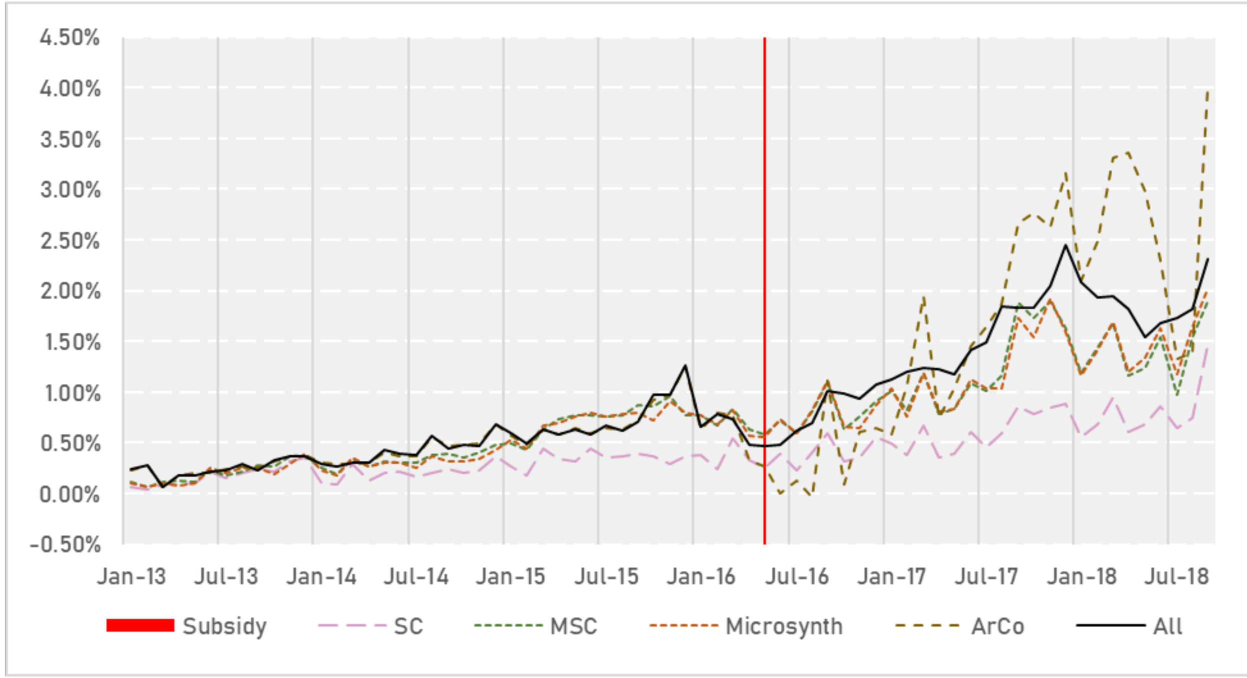
Note:

there are a plethora of graph types out there, we will cover the **most common**.

Graphs



1.40%
1.20%
1.00%
0.80%
0.60%
0.40%
0.20%
0.00%
Jan-13 Jul-13 Jan-14 Jul-14 Jan-15 Jul-15 Jan-16 Jul-16 Jan-17 Jul-17 Jan-18 Jul-18
Subsidy
SC
MSC
Microsynth
ArCo
PHEV



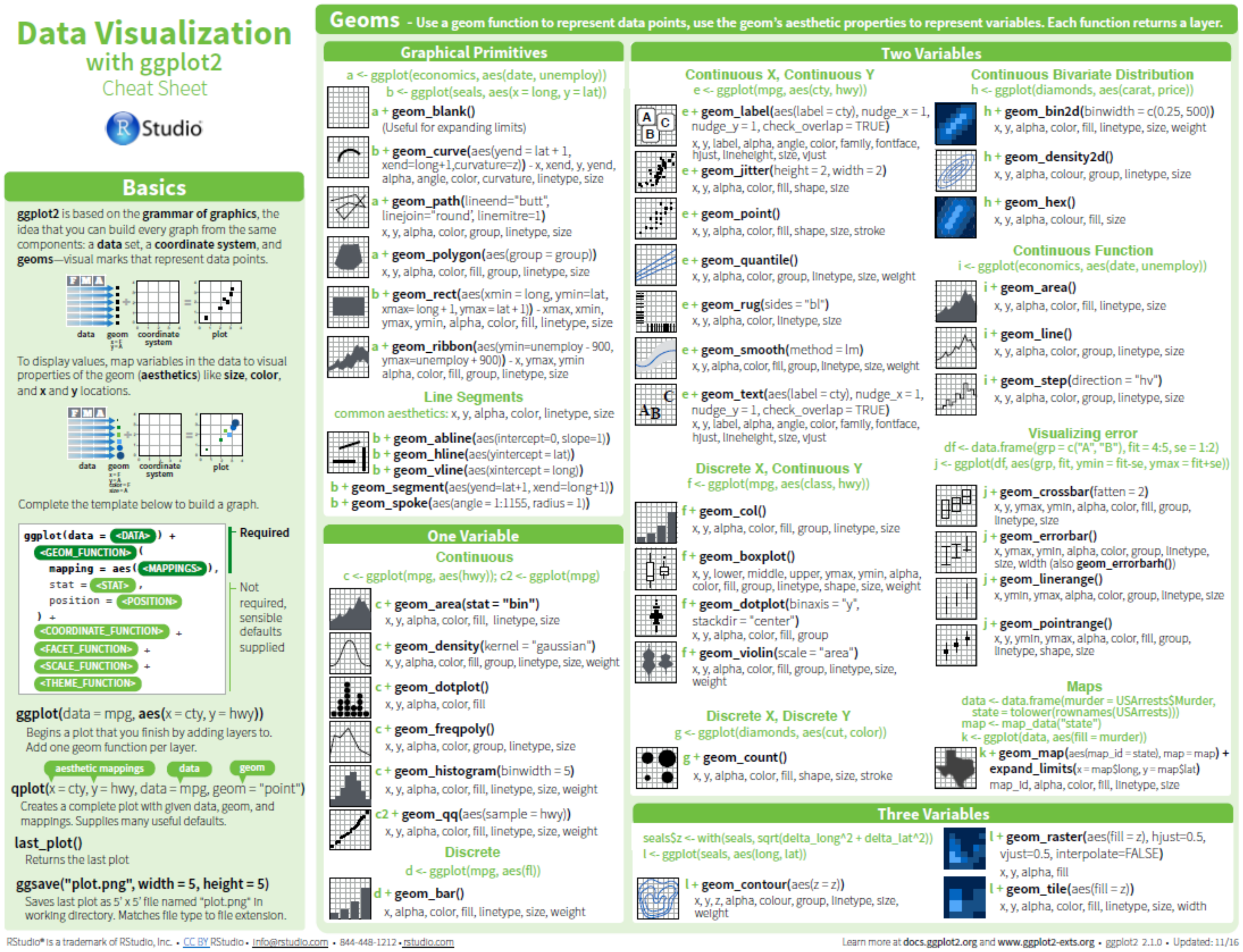
4.50%
4.00%
3.50%
3.00%
2.50%
2.00%
1.50%
1.00%
0.50%
0.00%
-0.50%
Jan-13 Jul-13 Jan-14 Jul-14 Jan-15 Jul-15 Jan-16 Jul-16 Jan-17 Jul-17 Jan-18 Jul-18
Subsidy
SC
MSC
Microsynth
ArCo
All

Environmentalist
Oil Worker
Flight Attendant
Pilot
Bartender
Beer Wholesaler
Librarian
Logger
Taxi Driver
Truck Driver
Innkeeper
Motel Owner
Chairwoman
Chairman
Pediatrician
Urologist
Sculptor
Plastic Surgeon
Floral Designer
Exterminator
Yoga Instructor
Car Salesman
Architect
Home Builder
Bookseller
Pawnbroker
Episcopal Priest
Catholic Priest
Carpenter
Plumber
Midwife
Surgeon
Park Ranger
Sheriff
Gardener
Farmer
Chef
Cattle Feeder
Comedian
Talk Show Host
Professional Poker Player
Insurance Agent
Union Organizer
Business Owner
Environmental Scientist
Petroleum Geologist
Psychiatrist
Neurosurgeon
Retail Trade
Garden & Landscape
Medical Services
Personal Services
Sports & Games
Religion
Civil Service
Public Relations
Foodservice
Engineering
Publishing
Legal
Missionary
Catholic Priest
Pastor
Reverend
Minister
Priest
Chaplain
Bishop
Clergy
Rabbi
Episcopal Priest

Plotting with ggplot2

The **concept** behind **ggplot2** divides plot into three (even more) different **fundamental** parts:
Plot = data + Aesthetics + Geometry

- **data** is a data frame
- **Aesthetics** is used to indicate x and y variables. It can also be used to control the **color**, the **size** or the **shape** of points, the height of bars, etc…..
- **Geometry** defines the type of graphics (**histogram**, **box plot**, **line plot**, **density plot**, **dot plot**, ….)
- All the rest….


Data Visualization with ggplot2 Cheat Sheet

RStudio

Basics

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data** set, a **coordinate system**, and **geoms**—visual marks that represent data points.

To display values, map variables in the data to visual properties of the geom (**aesthetics**) like **size**, **color**, and **x** and **y** locations.

Complete the template below to build a graph.

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(
    mapping = aes(<MAPPINGS>),
    stat = <STAT>,
    position = <POSITION>
  ) +
  <COORDINATE_FUNCTION> +
  <FACET_FUNCTION> +
  <SCALE_FUNCTION> +
  <THEME_FUNCTION>
```

Required. Not required, sensible defaults supplied

ggplot(data = mpg, **aes**(x = cty, y = hwy))
Begins a plot that you finish by adding layers to. Add one geom function per layer.

qplot(x = cty, y = hwy, data = mpg, geom = "point")
Creates a complete plot with given data, geom, and mappings. Supplies many useful defaults.

last_plot()
Returns the last plot

ggsave("plot.png", width = 5, height = 5)
Saves last plot as 5' x 5' file named "plot.png" in working directory. Matches file type to file extension.

Geoms - Use a geom function to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

Graphical Primitives
a <- ggplot(economics, aes(date, unemploy))
b <- ggplot(seals, aes(x = long, y = lat))
a + **geom_blank()** (Useful for expanding limits)
b + **geom_curve**(aes(yend = lat + 1, xend=long+1,curvature=z)) - x, xend, y, yend, alpha, angle, color, curvature, linetype, size
a + **geom_path**(lineend="butt", linejoin="round", linemitre=1) x, y, alpha, color, group, linetype, size
a + **geom_polygon**(aes(group = group)) x, y, alpha, color, fill, group, linetype, size
b + **geom_rect**(aes(xmin = long, ymin=lat, xmax= long + 1, ymax = lat + 1)) - xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size
a + **geom_ribbon**(aes(ymin=unemploy - 900, ymax=unemploy + 900)) - x, ymax, ymin alpha, color, fill, group, linetype, size

Line Segments
common aesthetics: x, y, alpha, color, linetype, size
b + **geom_abline**(aes(intercept=0, slope=1))
b + **geom_hline**(aes(yintercept = lat))
b + **geom_vline**(aes(xintercept = long))
b + **geom_segment**(aes(yend=lat+1, xend=long+1))
b + **geom_spoke**(aes(angle = 1:1155, radius = 1))

One Variable
Continuous
c <- ggplot(mpg, aes(hwy)); c2 <- ggplot(mpg)
c + **geom_area(stat = "bin")** x, y, alpha, color, fill, linetype, size
c + **geom_density**(kernel = "gaussian") x, y, alpha, color, fill, group, linetype, size, weight
c + **geom_dotplot()** x, y, alpha, color, fill
c + **geom_freqpoly()** x, y, alpha, color, group, linetype, size
c + **geom_histogram**(binwidth = 5) x, y, alpha, color, fill, linetype, size, weight
c2 + **geom_qq**(aes(sample = hwy)) x, y, alpha, color, fill, linetype, size, weight

Discrete
d <- ggplot(mpg, aes(fl))
d + **geom_bar()** x, alpha, color, fill, linetype, size, weight

Two Variables
Continuous X, Continuous Y
e <- ggplot(mpg, aes(cty, hwy))
e + **geom_label**(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE) x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust
e + **geom_jitter**(height = 2, width = 2) x, y, alpha, color, fill, shape, size
e + **geom_point()** x, y, alpha, color, fill, shape, size, stroke
e + **geom_quantile()** x, y, alpha, color, group, linetype, size, weight
e + **geom_rug**(sides = "bl") x, y, alpha, color, linetype, size
e + **geom_smooth**(method = lm) x, y, alpha, color, fill, group, linetype, size, weight
e + **geom_text**(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE) x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

Discrete X, Continuous Y
f <- ggplot(mpg, aes(class, hwy))
f + **geom_col()** x, y, alpha, color, fill, group, linetype, size
f + **geom_boxplot()** x, y, lower, middle, upper, ymax, ymin, alpha, color, fill, group, linetype, shape, size, weight
f + **geom_dotplot**(binaxis = "y", stackdir = "center") x, y, alpha, color, fill, group
f + **geom_violin**(scale = "area") x, y, alpha, color, fill, group, linetype, size, weight

Discrete X, Discrete Y
g <- ggplot(diamonds, aes(cut, color))
g + **geom_count()** x, y, alpha, color, fill, shape, size, stroke

Continuous Bivariate Distribution
h <- ggplot(diamonds, aes(carat, price))
h + **geom_bin2d**(binwidth = c(0.25, 500)) x, y, alpha, color, fill, linetype, size, weight
h + **geom_density2d()** x, y, alpha, colour, group, linetype, size
h + **geom_hex()** x, y, alpha, colour, fill, size

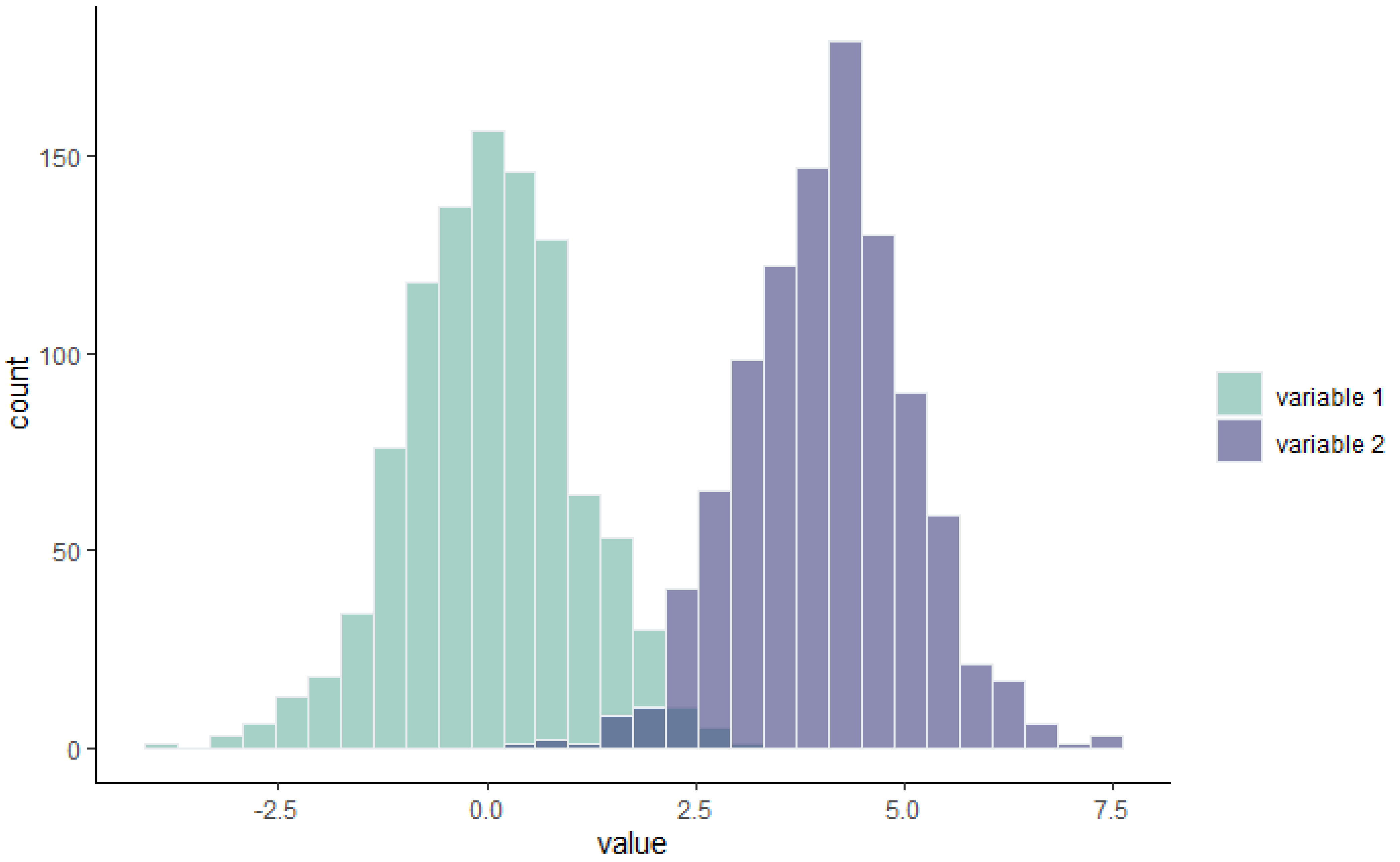
Continuous Function
i <- ggplot(economics, aes(date, unemploy))
i + **geom_area()** x, y, alpha, color, fill, linetype, size
i + **geom_line()** x, y, alpha, color, group, linetype, size
i + **geom_step**(direction = "hv") x, y, alpha, color, group, linetype, size

Visualizing error
df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)
j <- ggplot(df, aes(grp, fit, ymin = fit-se, ymax = fit+se))
j + **geom_crossbar**(fatten = 2) x, y, ymax, ymin, alpha, color, fill, group, linetype, size
j + **geom_errorbar()** x, ymax, ymin, alpha, color, group, linetype, size, width (also **geom_errorbarh()**)
j + **geom_linerange()** x, ymin, ymax, alpha, color, group, linetype, size
j + **geom_pointrange()** x, y, ymin, ymax, alpha, color, fill, group, linetype, shape, size

Maps
data <- data.frame(murder = USArrests$Murder, state = tolower(rownames(USArrests)))
map <- map_data("state")
k <- ggplot(data, aes(fill = murder))
k + **geom_map**(aes(map_id = state), map = map) + **expand_limits**(x = map$long, y = map$lat) map_id, alpha, color, fill, linetype, size

Three Variables
seals$z <- with(seals, sqrt(delta_long^2 + delta_lat^2))
l <- ggplot(seals, aes(long, lat))
l + **geom_contour**(aes(z = z)) x, y, z, alpha, colour, group, linetype, size, weight
l + **geom_raster**(aes(fill = z), hjust=0.5, vjust=0.5, interpolate=FALSE) x, y, alpha, fill
l + **geom_tile**(aes(fill = z)) x, y, alpha, color, fill, linetype, size, width

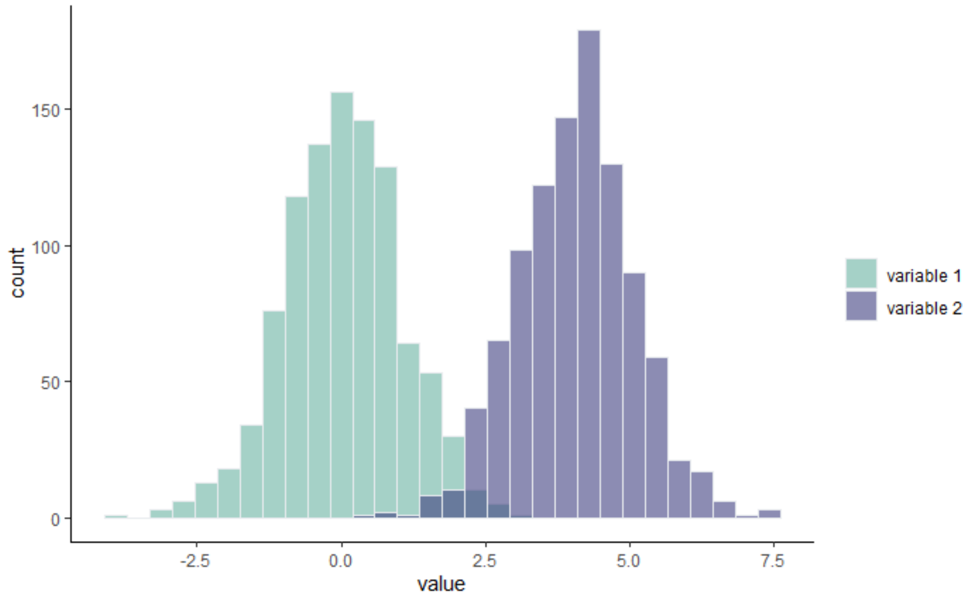
Learn more at docs.ggplot2.org and www.ggplot2-exts.org • ggplot2 2.1.0 • Updated: 11/16


150
100
50
0
count
-2.5
0.0
2.5
5.0
7.5
value
variable 1
variable 2

Example

```
data <- data.frame(
        type = c( rep("variable 1", 1000), rep("variable 2", 1000) ),
        value = c( rnorm(1000), rnorm(1000, mean=4) ) ) # Represent it

ggplot(data, aes(x=value, fill=type)) +
        geom_histogram( color="#e9ecef", alpha=0.6, position = 'identity') +
        scale_fill_manual(values=c("#69b3a2", "#404080")) + theme_classic()
        + labs(fill="")
```

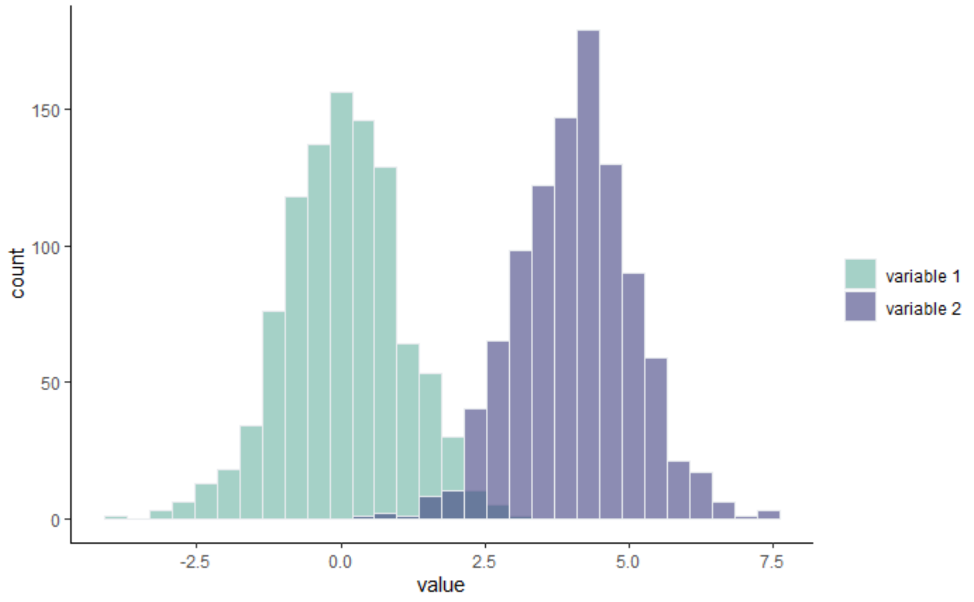


1. Generate a dataframe with two variables and 1000 observations
2. Generate random values based on a normal distribution

Example

```
data <- data.frame(
            type = c( rep("variable 1", 1000), rep("variable 2", 1000) ),
            value = c( rnorm(1000), rnorm(1000, mean=4) ) ) # Represent it

ggplot(data, aes(x=value, fill=type)) +
            geom_histogram( color="#e9ecef", alpha=0.6, position = 'identity') +
            scale_fill_manual(values=c("#69b3a2", "#404080")) + theme_classic()
            + labs(fill="")
```

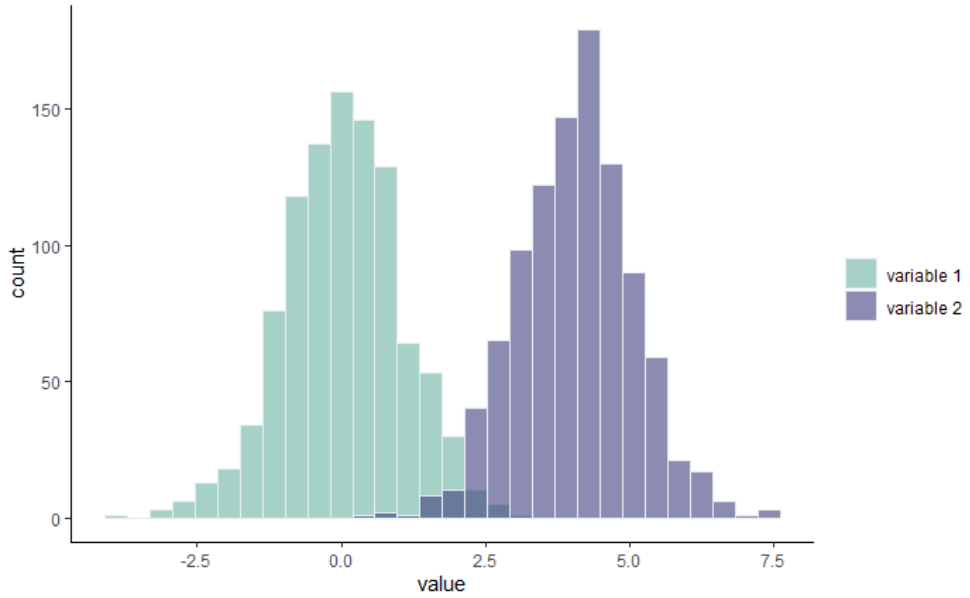


1. Generate a dataframe with two variables and 1000 observations
2. Generate random values based on a normal distribution
3. Select the data and variables we want to plot

Example

```
data <- data.frame(
        type = c( rep("variable 1", 1000), rep("variable 2", 1000) ),
        value = c( rnorm(1000), rnorm(1000, mean=4) ) ) # Represent it

ggplot(data, aes(x=value, fill=type)) +
        geom_histogram( color="#e9ecef", alpha=0.6, position = 'identity') +
        scale_fill_manual(values=c("#69b3a2", "#404080")) + theme_classic()
        + labs(fill="")
```

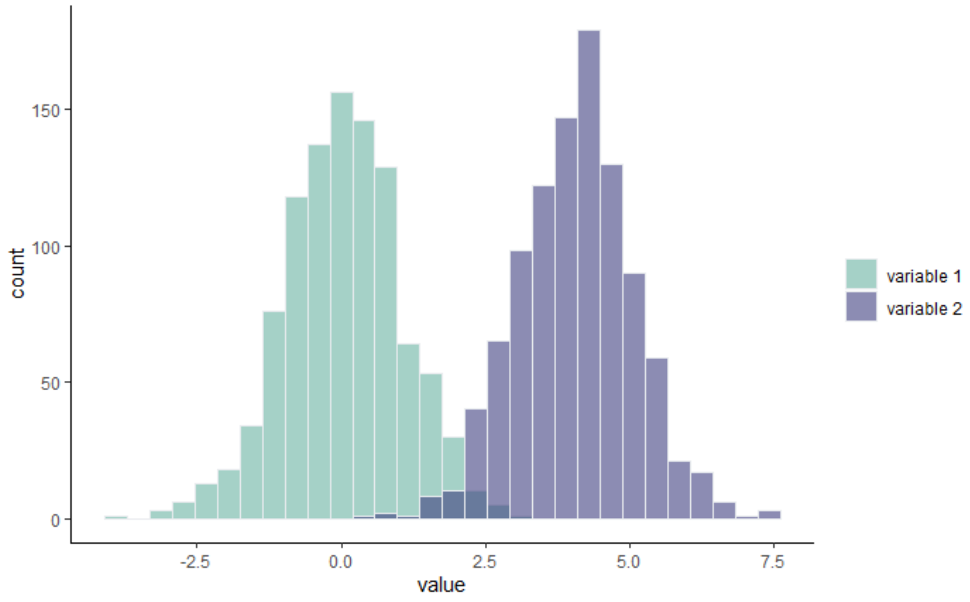


1. Generate a dataframe with two variables and 1000 observations
2. Generate random values based on a normal distribution
3. Select the data and variables we want to plot
4. Select the type of visualization, identify, color sets the lines, alpha for transparency and identify overlaps the bar

Example

```
data <- data.frame(
          type = c( rep("variable 1", 1000), rep("variable 2", 1000) ),
          value = c( rnorm(1000), rnorm(1000, mean=4) ) ) # Represent it

ggplot(data, aes(x=value, fill=type)) +
          geom_histogram( color="#e9ecef", alpha=0.6, position = 'identity') +
          scale_fill_manual(values=c("#69b3a2", "#404080")) + theme_classic()
          + labs(fill="")
```

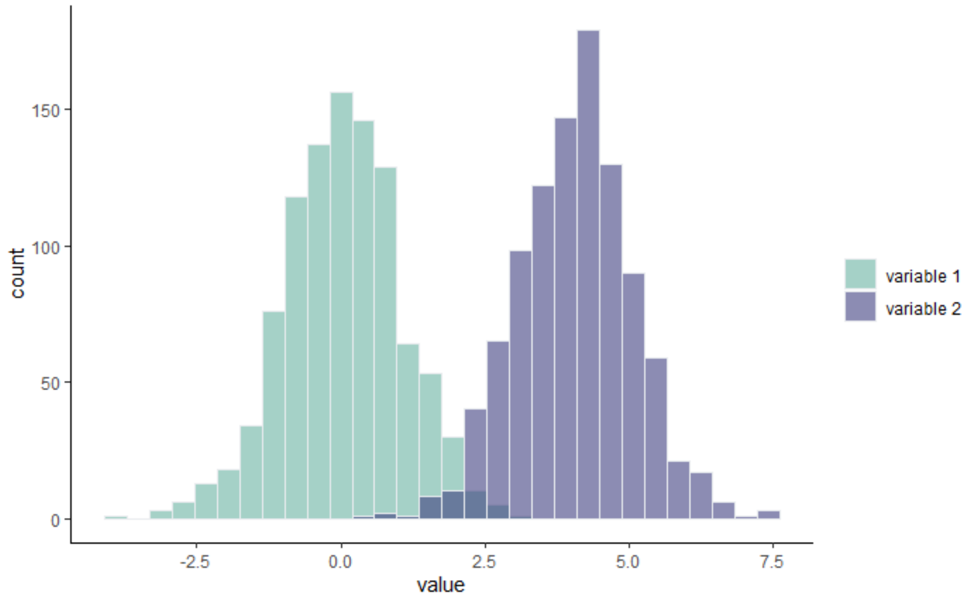


1. Generate a dataframe with two variables and 1000 observations
2. Generate random values based on a normal distribution
3. Select the data and variables we want to plot
4. Select the type of visualization, identify, color sets the lines, alpha for transparency and identify overlaps the bar
5. We fill each histogram with a different colour

Example

```
data <- data.frame(
          type = c( rep("variable 1", 1000), rep("variable 2", 1000) ),
          value = c( rnorm(1000), rnorm(1000, mean=4) ) ) # Represent it

ggplot(data, aes(x=value, fill=type)) +
          geom_histogram( color="#e9ecef", alpha=0.6, position = 'identity') +
          scale_fill_manual(values=c("#69b3a2", "#404080")) + theme_classic()
          + labs(fill="")
```

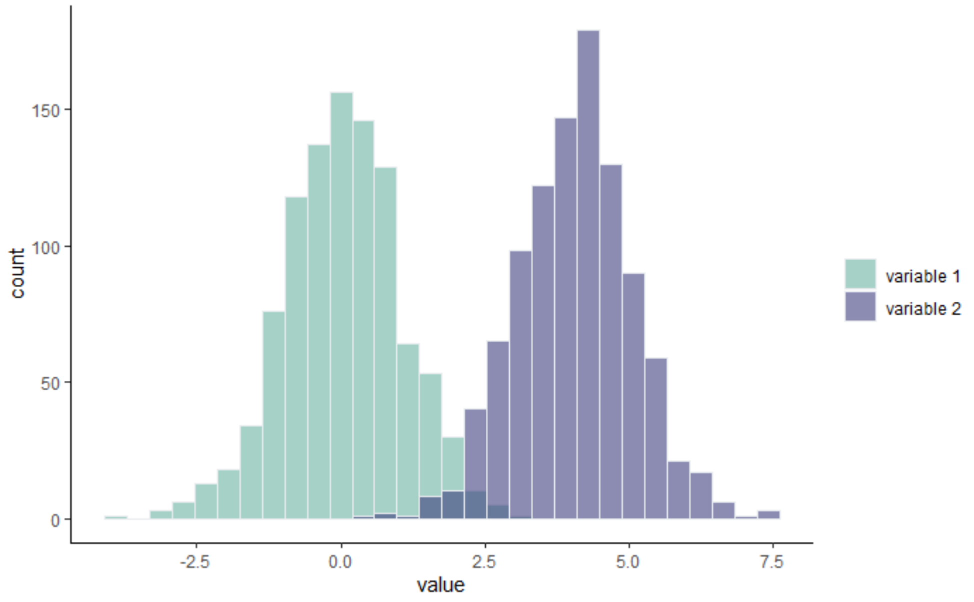


1. Generate a dataframe with two variables and 1000 observations
2. Generate random values based on a normal distribution
3. Select the data and variables we want to plot
4. Select the type of visualization, identify, color sets the lines, alpha for transparency and identify overlaps the bar
5. We fill each histogram with a different colour
6. Choose a preset theme

Example

```r
data <- data.frame(
          type = c( rep("variable 1", 1000), rep("variable 2", 1000) ),
          value = c( rnorm(1000), rnorm(1000, mean=4) ) ) # Represent it

ggplot(data, aes(x=value, fill=type)) +
          geom_histogram( color="#e9ecef", alpha=0.6, position = 'identity') +
          scale_fill_manual(values=c("#69b3a2", "#404080")) + theme_classic()
          + labs(fill="")
```



1. Generate a dataframe with two variables and 1000 observations
2. Generate random values based on a normal distribution
3. Select the data and variables we want to plot
4. Select the type of visualization, identify, color sets the lines, alpha for transparency and identify overlaps the bar
5. We fill each histogram with a different colour
6. Choose a preset theme
7. Names for labs added

Data Visualization

Prof. Dr. Javier Valdes
Javier.valdes@th-deg.de

IAI - Institut für Angewandte Informatik
Institute for Applied Informatics

Technische Hochschule Deggendorf
Technologie Campus Freyung
Grafenauer Str. 22, D- 94078 Freyung, Germany

Tel.: +49 8551 91764 40
Fax: +49 8551 91764 69