

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

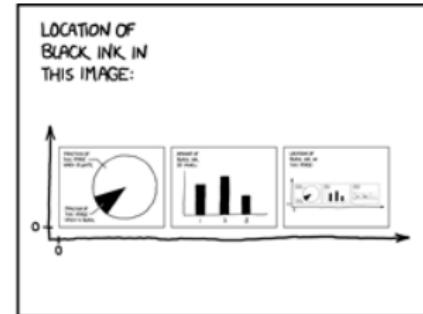
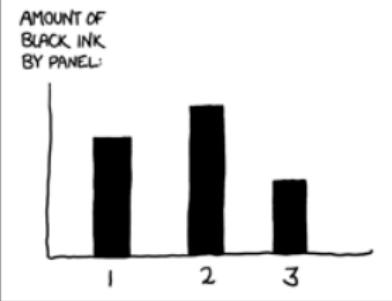
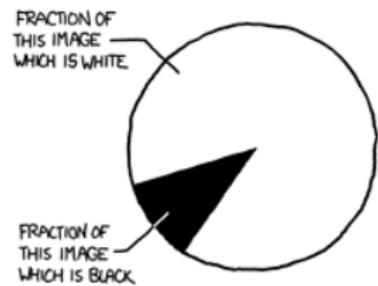
Rogue's
Gallery

References

Data Visualization

Dennis Wylie, UT Bioinformatics Consulting Group

April 15, 2017



Types of visualization [Hegarty (2011)]

Data Visualization

xkcd

History

Principles

Grammar of graphics

Minard

Composite graphics

Rogue's Gallery

References

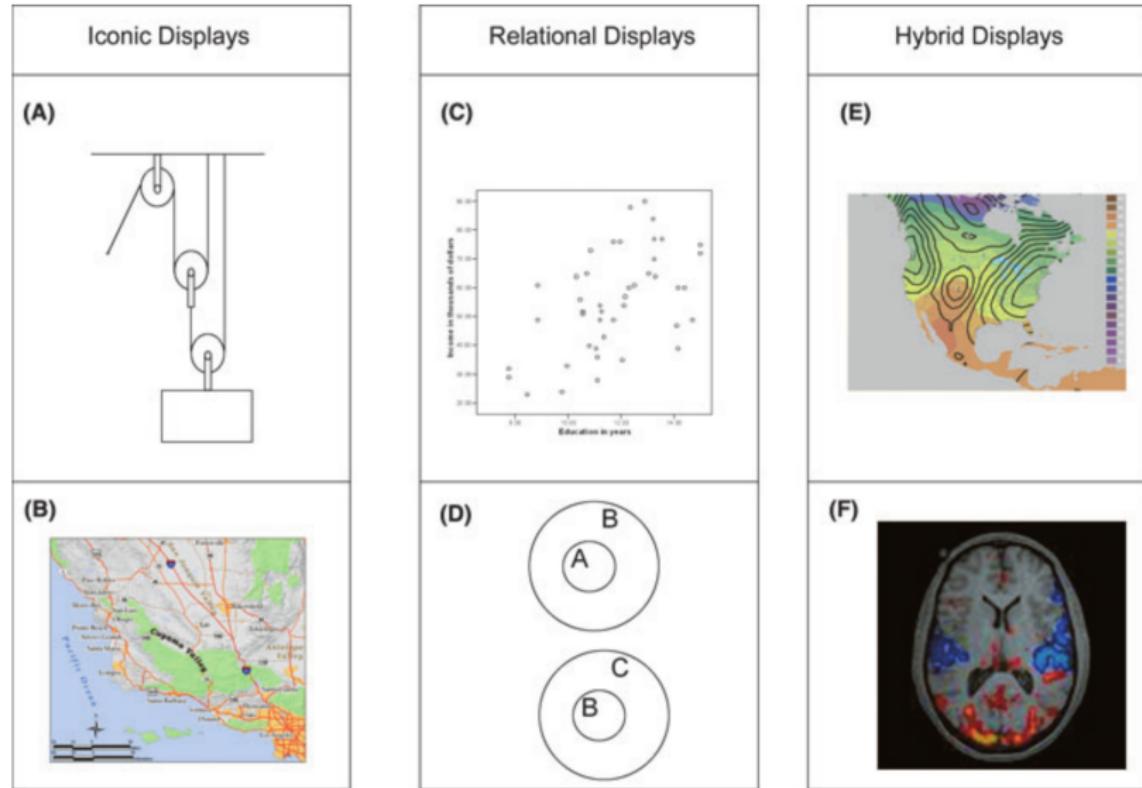


Fig. 1. Examples of iconic, relational, and hybrid displays.

History of data visualization [Friendly (2008)]

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

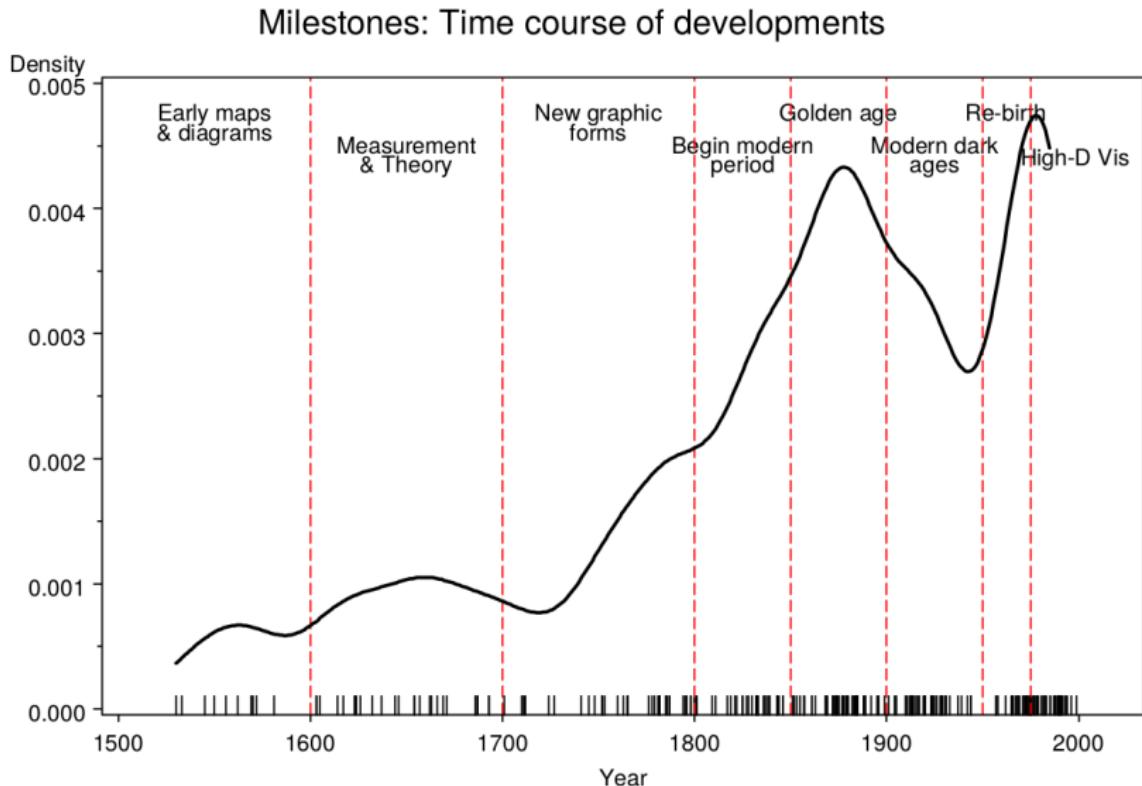


Figure 1: The time distribution of events considered milestones in the history of data visualization, shown by a rug plot and density estimate.

Playfair

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

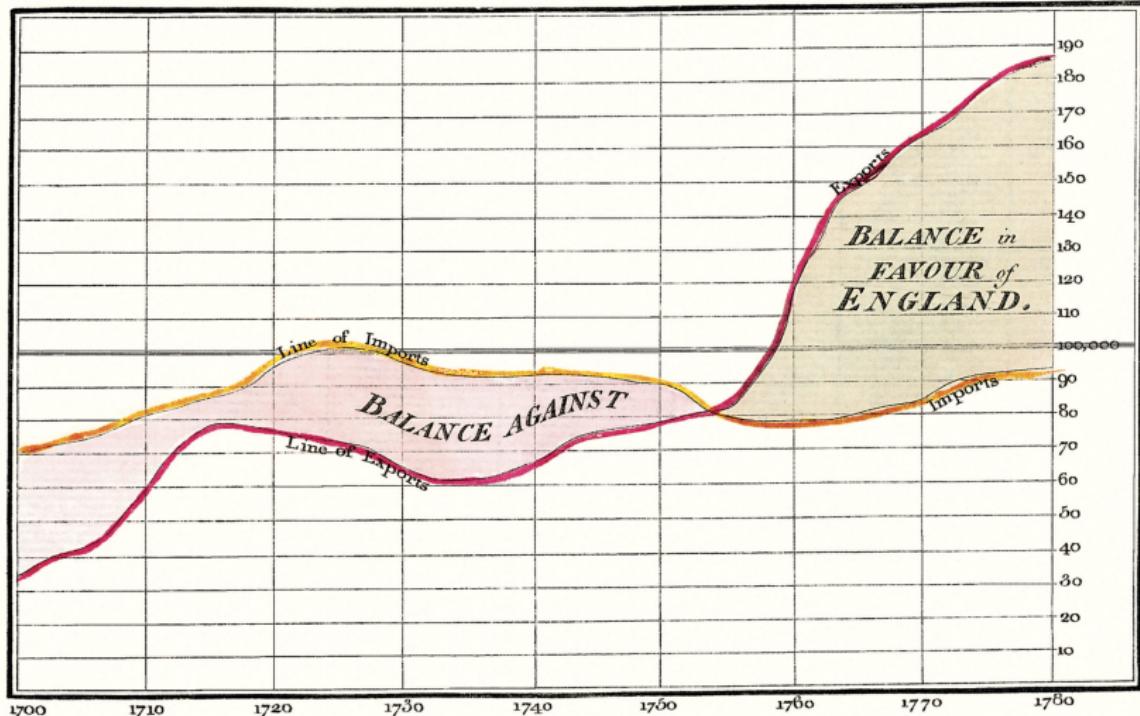
Minard

Composite
graphics

Rogue's
Gallery

References

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



The Bottom line is divided into Years, the Right hand line into £10,000 each.

Published as the Act directs, 1st May 1786, by W^m Playfair

Noel script 352, Strand, London.

Minard

Data
Visualization

xkcd

History

Principles

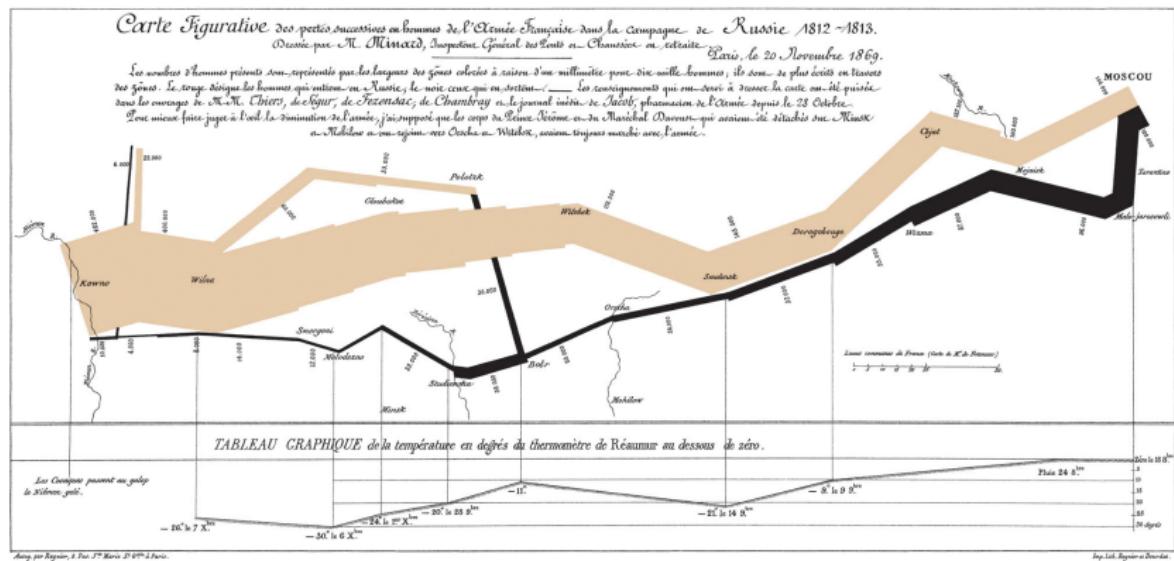
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



Nightingale

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

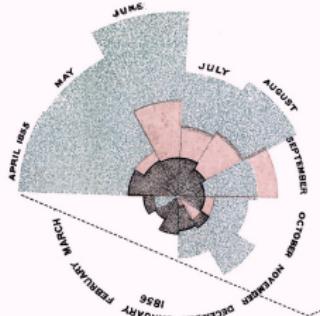
Minard

Composite
graphics

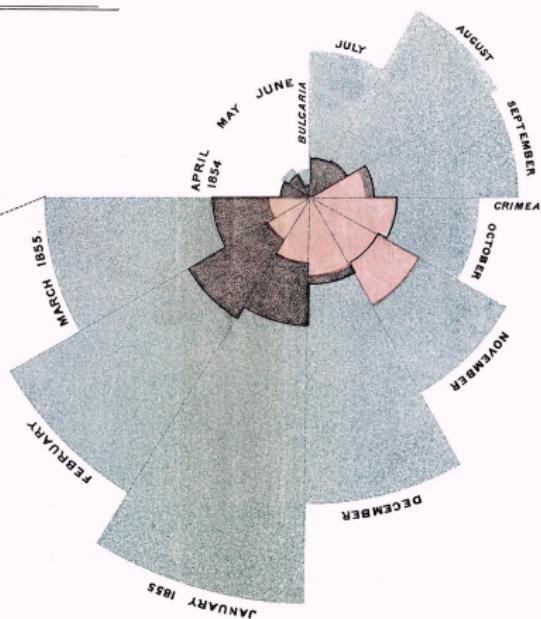
Rogue's
Gallery

References

2.
APRIL 1855 to MARCH 1856.
DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.



1.
APRIL 1854 to MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes. The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire area may be compared by following the blue, the red & the black lines enclosing them.

Why visualize?

Data
Visualization

- ▶ Summarize data

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

Why visualize?

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

- ▶ Summarize data
- ▶ Illustrate relationships

Why visualize?

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

- ▶ Summarize data
- ▶ Illustrate relationships
- ▶ Pattern recognition by human visual perception

Why visualize?

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

- ▶ Summarize data
- ▶ Illustrate relationships
- ▶ Pattern recognition by human visual perception
 - ▶ map properties to “aesthetics”

Why visualize?

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

- ▶ Summarize data
- ▶ Illustrate relationships
- ▶ Pattern recognition by human visual perception
 - ▶ map properties to “aesthetics”

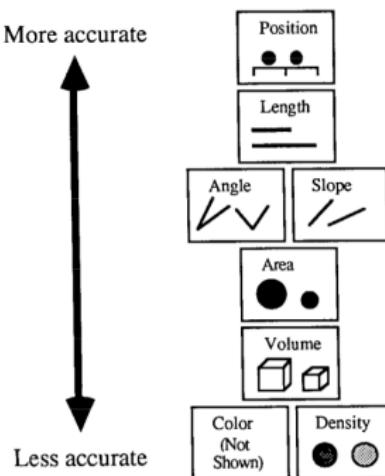


Fig. 14. Accuracy ranking of quantitative perceptual tasks. Higher tasks are accomplished more accurately than lower tasks. Cleveland and McGill empirically verified the basic properties of this ranking.

Taken from Mackinlay (1986)

Natural example: frequency mapped to angle

Data
Visualization

xkcd

History

Principles

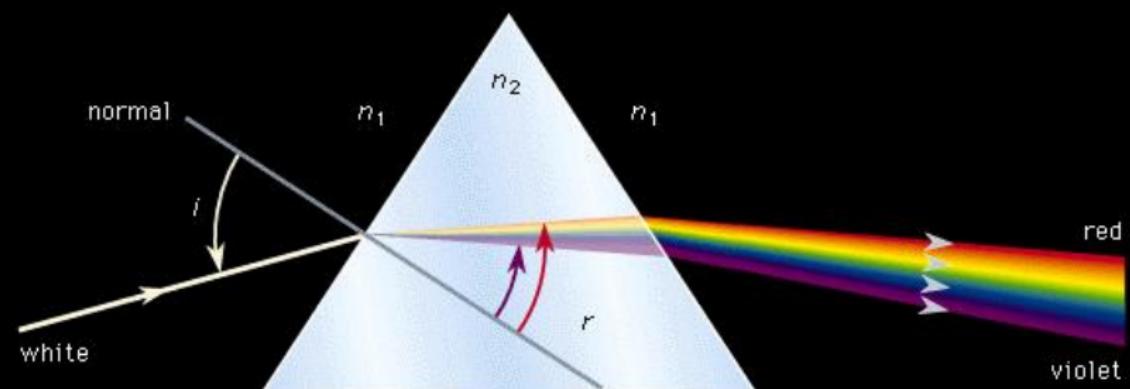
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



The angles i and r that the rays make with the normal are the angles of incidence and refraction. Because n_2 depends upon wavelength, the incident white ray separates into its constituent colours upon refraction, with deviation of the red ray the least and the violet ray the most.

Experimental example: size mapped to position

Data
Visualization

xkcd

History

Principles

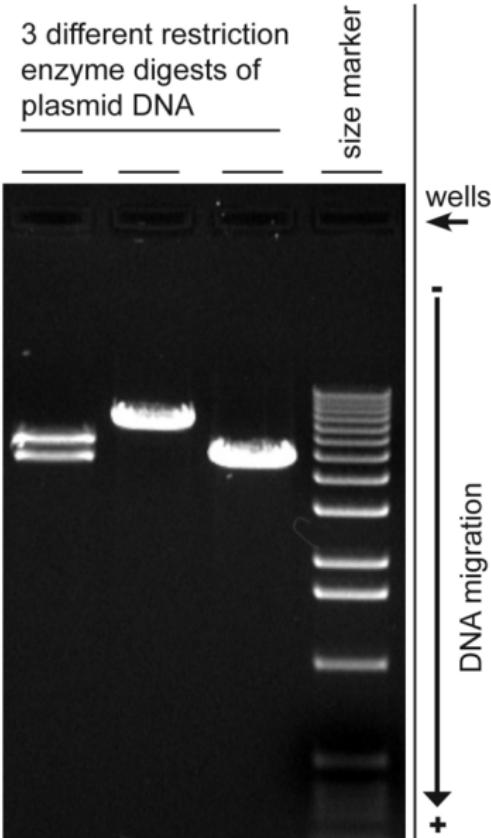
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



- ▶ Discrete experiment/condition mapped to horizontal position
- ▶ Continuous(ish) molecular size mapped to vertical position
- ▶ Continuous(ish) copy number mapped to brightness

Choosing the right aesthetic

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

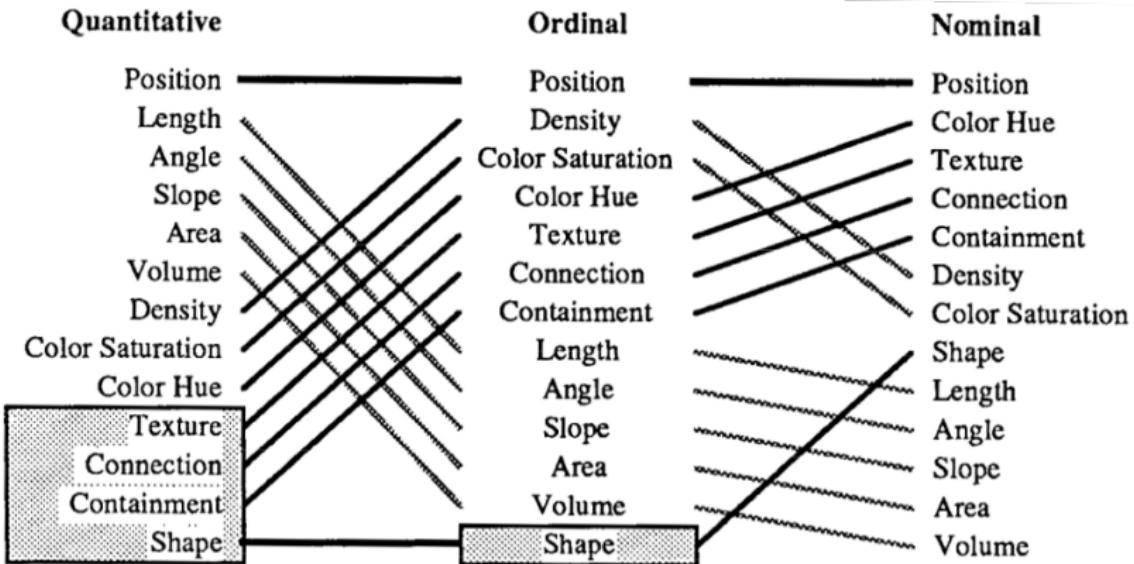


Fig. 15. Ranking of perceptual tasks. The tasks shown in the gray boxes are not relevant to these types of data.

Taken from Mackinlay (1986).

"Lie factor" from improper area encoding

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

	1964	1975	1990
27 %		16.0 %	12.0 %



Taken from Tufts (1982)

Neuron subtypes and GSE75386

Data
Visualization

xkcd

History

Principles

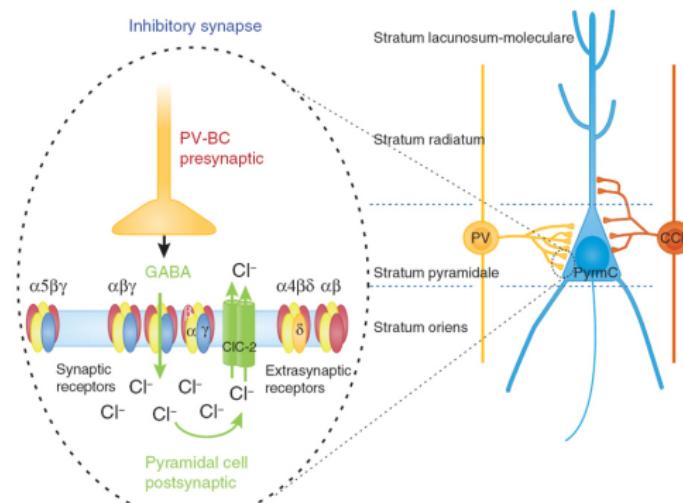
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

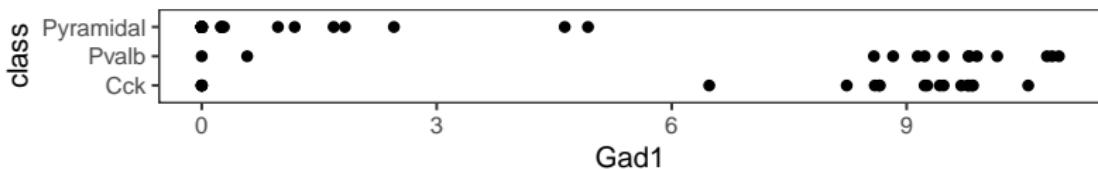


GSE75386 Single cell RNASeq of electrophysiologically characterized neurons of the hippocampus

Organism *Mus musculus*

Design 15 tissue samples + 93 single-cell samples, including 58 *Cck*, *Pvalb*, and Pyramidal neurons.

GSE75386 stripchart example



Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

GSE75386 stripchart example

Data
Visualization

xkcd

History

Principles

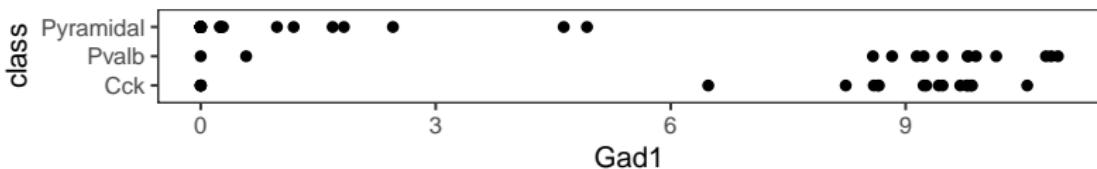
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggstrip = ggplot(  
  data = gse75386,  
  mapping = aes(  
    x = Gad1,  
    y = class  
  )  
)  
ggstrip = ggstrip + geom_point()  
print(ggstrip)
```

GSE75386 stripchart example

Data
Visualization

xkcd

History

Principles

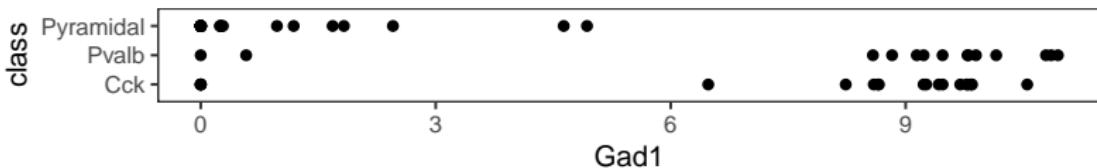
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggstrip = ggplot(  
  data = gse75386,  
  mapping = aes(  
    x = Gad1,  
    y = class  
  )  
)  
ggstrip = ggstrip + geom_point()  
print(ggstrip)
```

What is “geom_point”?

GSE75386 overplotted bars

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggbar = ggplot(gse75386, aes(x=class, y=Gad1))
ggbar = ggbar + geom_bar(alpha=0.1,
                         position='identity', stat='identity')
ggbar = ggbar + coord_flip()
print(ggbar)
```

GSE75386 mean bars + SE lines

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
gse75386stats = gse75386 %>%
  group_by(class) %>%
  summarize(
    `Gad1 (Mean)` = mean(Gad1),
    SE = sd(Gad1) / sqrt(length(Gad1))
  )
ggbarse = ggplot(gse75386stats, aes(x=class, y=`Gad1 (Mean)`))
ggbarse = ggbarse + geom_bar(alpha=0.6, stat='identity')
ggbarse = ggbarse + geom_errorbar(aes(ymax=`Gad1 (Mean)` - SE,
                                       ymax=`Gad1 (Mean)` + SE),
                                    width=0)
ggbarse = ggbarse + coord_flip()
print(ggbarse)
```

Layered grammar of graphics [Wickham (2010)]

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

layer = **data** +
statistical transformation +
aesthetic +
geometric object

GSE75386 boxplot + stripchart

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggbox = ggplot(gse75386, aes(x=class, y=Gad1))
ggbox = ggbox + geom_boxplot(stat='boxplot',
                             outlier.size=0)
ggbox = ggbox + geom_point(alpha=0.5)
ggbox = ggbox + coord_flip()
```

GSE75386 boxplot + stripchart

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggbox = ggplot(gse75386, aes(x=class, y=Gad1))
ggbox = ggbox + geom_boxplot(stat='boxplot',
                             outlier.size=0)
ggbox = ggbox + geom_point(alpha=0.5)
ggbox = ggbox + coord_flip()
```

... some statistical transformations are built in to ggplot.

GSE75386 scatterplot

Data
Visualization

xkcd

History

Principles

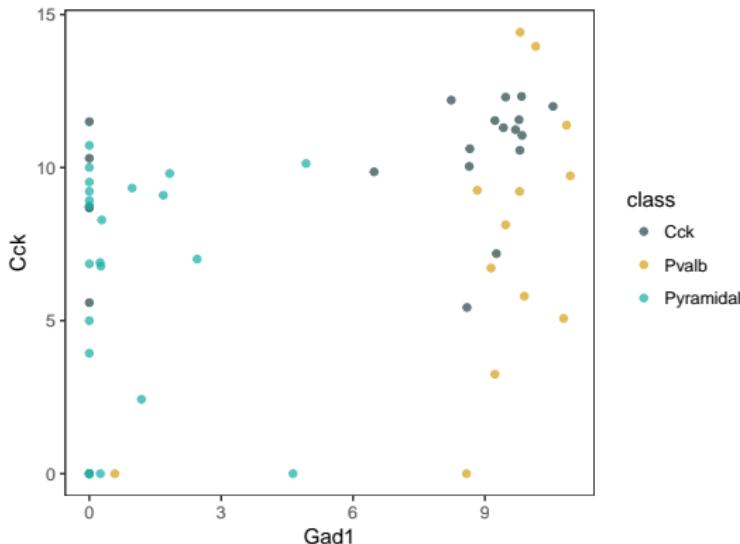
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggscat = ggplot(  
  gse75386,  
  aes(x=Gad1, y=Cck, color=class))  
ggscat = ggscat + geom_point(alpha=0.75)  
ggscat = ggscat + scale_color_manual(  
  values=c('darkslategray', 'goldenrod', 'lightseagreen'))
```

GSE75386 scatterplot

Data Visualization

xkcd

History

Principles

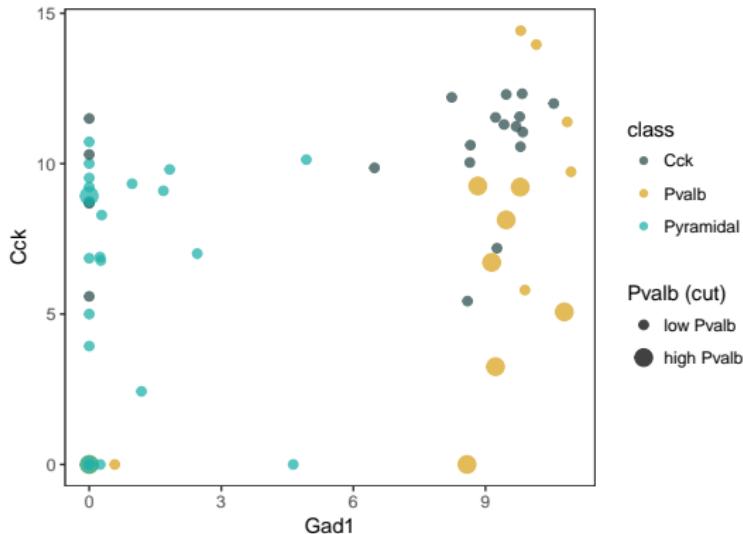
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



class

- Cck
- Pvalb
- Pyramidal

Pvalb (cut)

- low Pvalb
- high Pvalb

```
ggscat = ggplot(  
    gse75386,  
    aes(x=Gad1, y=Cck, color=class, size=`Pvalb (cut)`)  
)  
ggscat = ggscat + geom_point(alpha=0.75)  
ggscat = ggscat + scale_color_manual(  
    values=c('darkslategray', 'goldenrod', 'lightseagreen'))  
ggscat = ggscat + scale_size_manual(values=c(2, 4))
```

GSE75386 scatterplot + text layer

Data
Visualization

xkcd

History

Principles

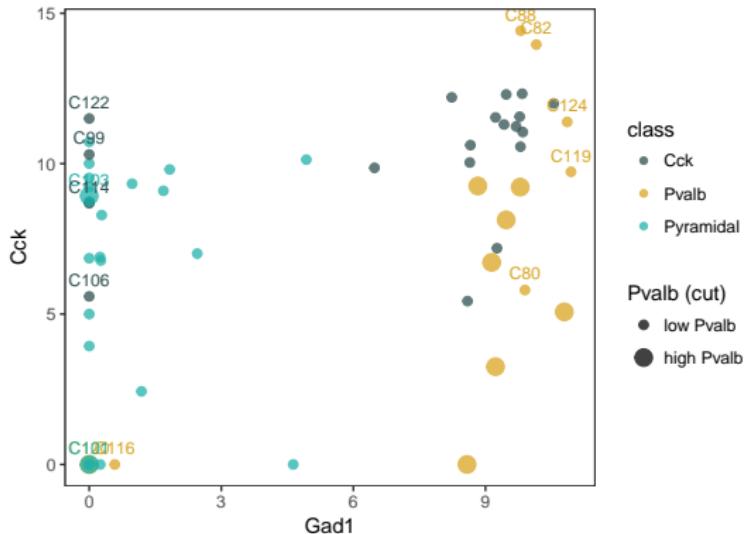
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
...  
ggscat = ggscat + geom_text_repel(  
  aes(label=odd),  
  size = 3,  
  show.legend = FALSE  
)
```

The best graph ever?

Data
Visualization

xkcd

History

Principles

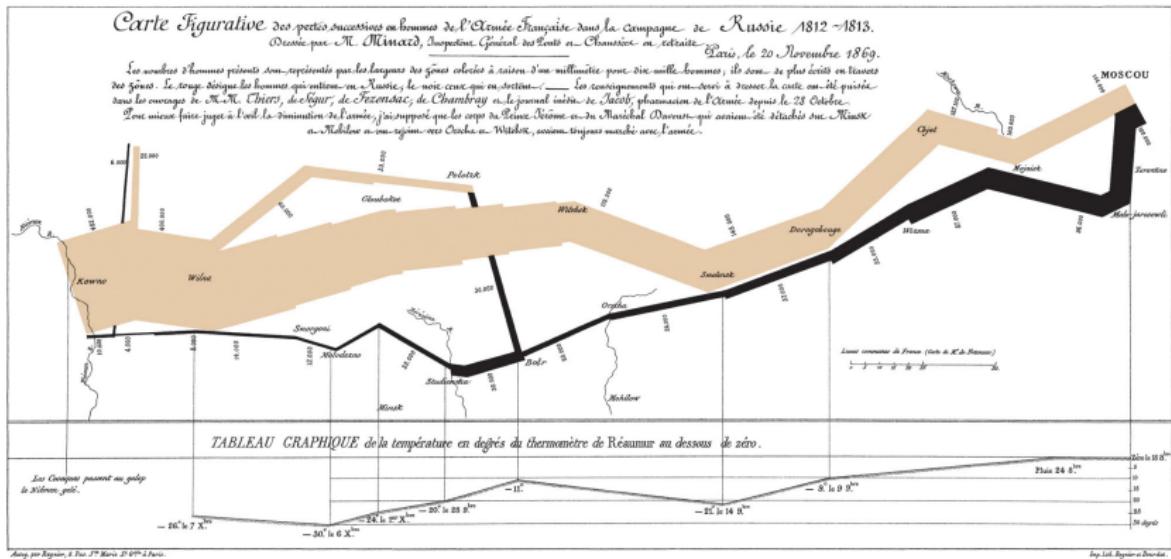
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



ggplotting Minard

Data Visualization

xkcd

History

Principles

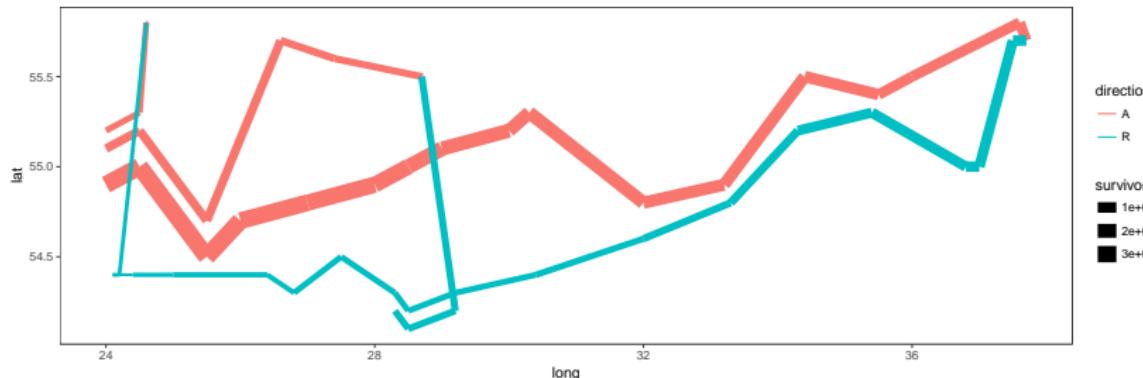
Grammar of graphics

Minard

Composite graphics

Rogue's Gallery

References



```
ggtroops = ggplot(troops, aes(long, lat))
ggtroops = ggtroops + geom_path(aes(
    size = survivors,
    color = direction,
    group = group
))
```

ggplotting Minard

Data
Visualization

xkcd

History

Principles

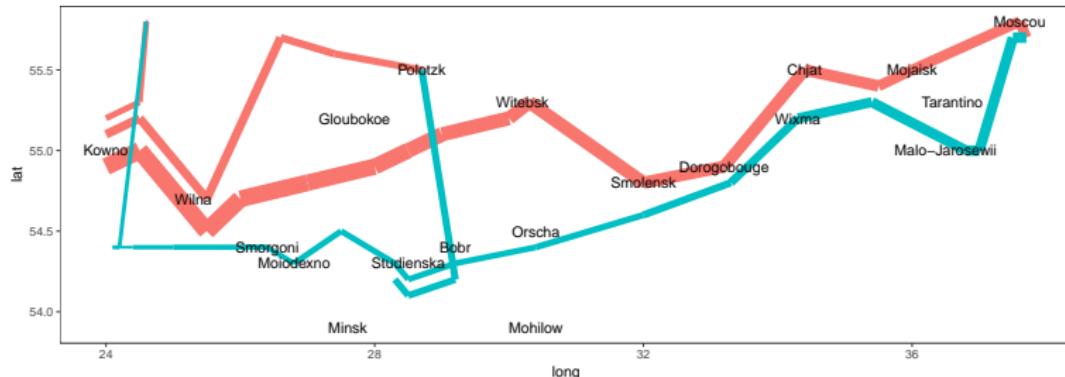
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggbboth = ggtroops + geom_text(  
    aes(label = city),  
    size = 4,  
    data = cities  
)
```

ggplotting Minard

Data
Visualization

xkcd

History

Principles

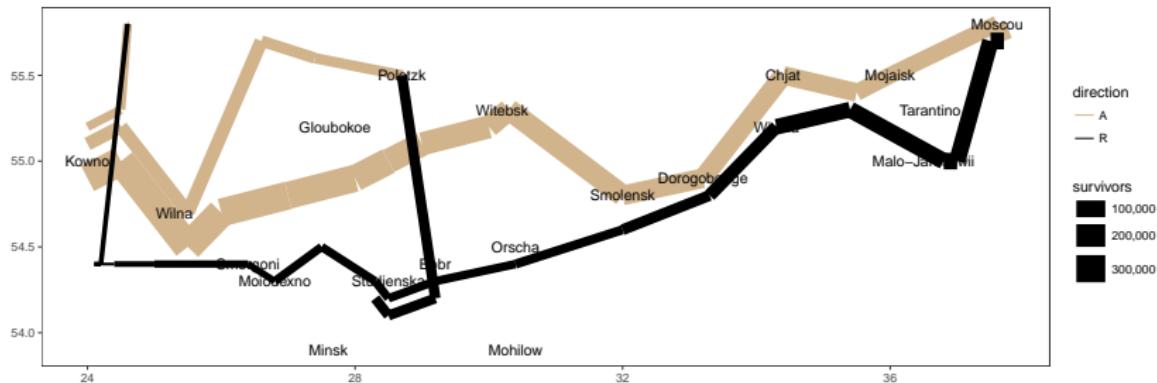
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggbboth = ggbboth + scale_size(  
    range = c(1, 10),  
    breaks = c(1, 2, 3) * 10^5,  
    labels = comma(c(1, 2, 3) * 10^5)  
)  
ggbboth = ggbboth + scale_color_manual(values = c("#d2b48c","black"))  
ggbboth = ggbboth + xlab(NULL) + ylab(NULL)
```

Small multiples and facetting

Data
Visualization

xkcd

History

Principles

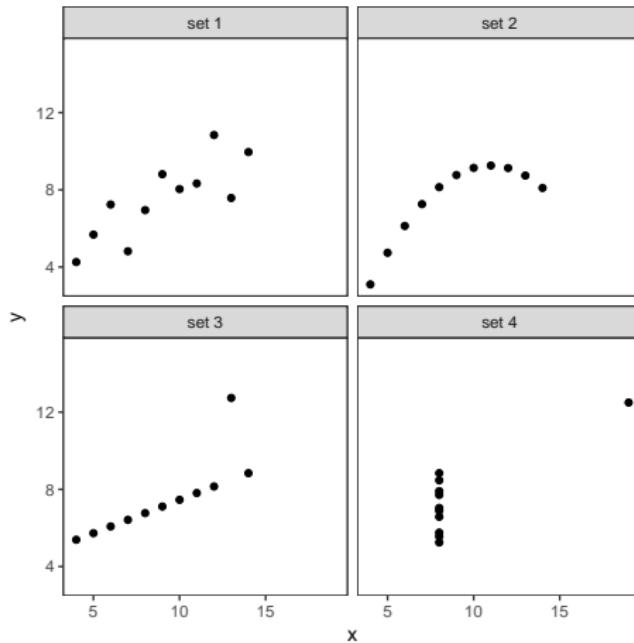
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



Taken from Anscombe (1973).

```
ggo = ggplot(anscombe, aes(x=x, y=y))
ggo = ggo + facet_wrap(~ set)
ggo = ggo + geom_point()
```

... with smoothed linear fit

Data Visualization

xkcd

History

Principles

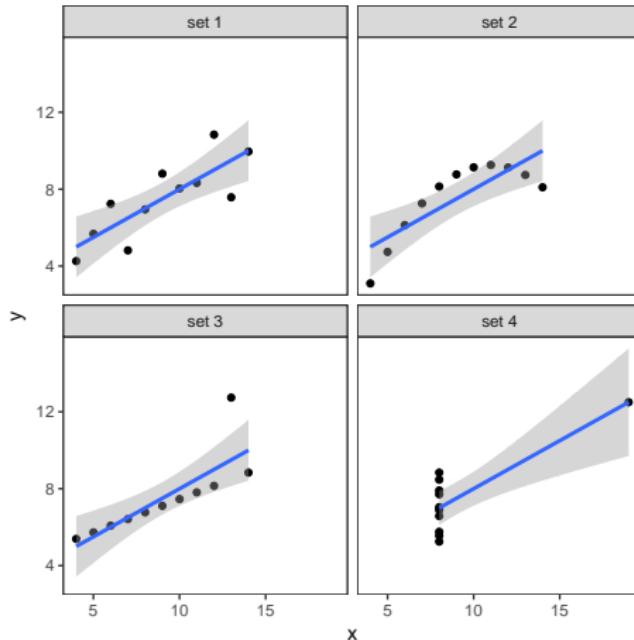
Grammar of graphics

Minard

Composite graphics

Rogue's Gallery

References



```
ggo = ggplot(anscombe, aes(x=x, y=y))
ggo = ggo + facet_wrap(~ set)
ggo = ggo + geom_point()

ggo = ggo + stat_smooth(method=lm)
```

... and text layer

Data Visualization

xkcd

History

Principles

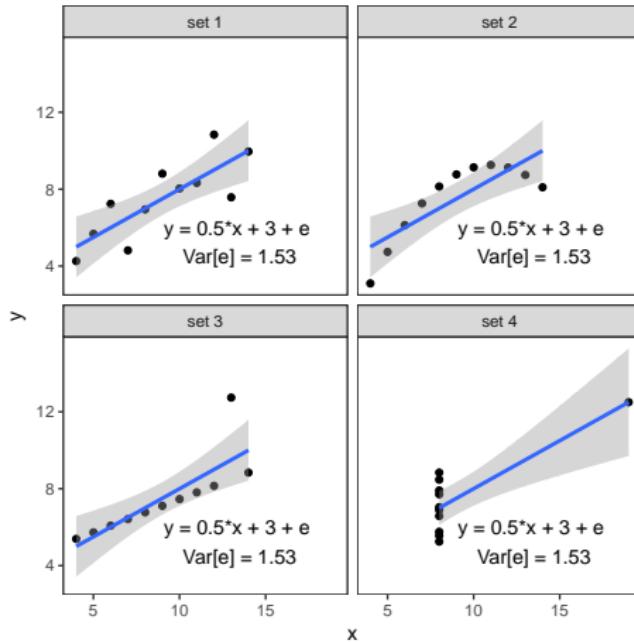
Grammar of graphics

Minard

Composite graphics

Rogue's Gallery

References



```
ggo = ggplot(anscombe, aes(x=x, y=y))
ggo = ggo + facet_wrap(~ set)
ggo = ggo + geom_point()
ggo = ggo + geom_text(aes(label=text), data=lmDescriptions(anscombe))
ggo = ggo + stat_smooth(method=lm)
```

Other smoothing methods exist as well

Data
Visualization

xkcd

History

Principles

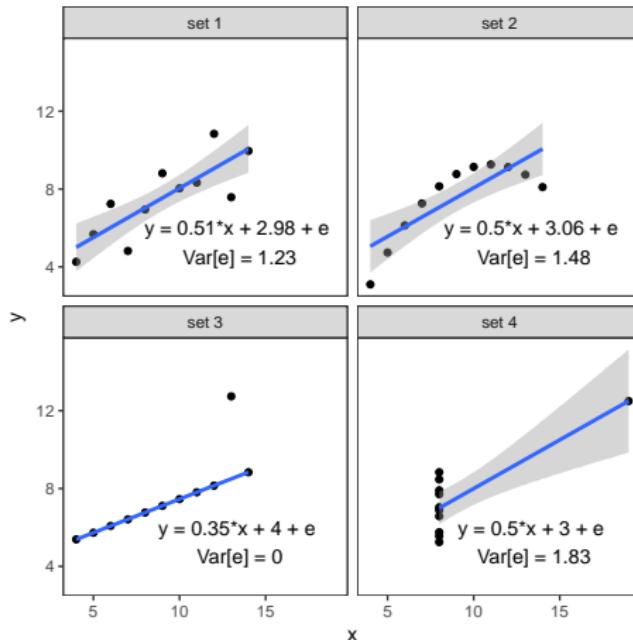
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggo = ggplot(anscombe, aes(x=x, y=y))
ggo = ggo + facet_wrap(~ set)
ggo = ggo + geom_point()
ggo = ggo + geom_text(aes(label=text), data=lmDescriptions(anscombe))
ggo = ggo + stat_smooth(method=rlm)
```

Local regression

Data
Visualization

xkcd

History

Principles

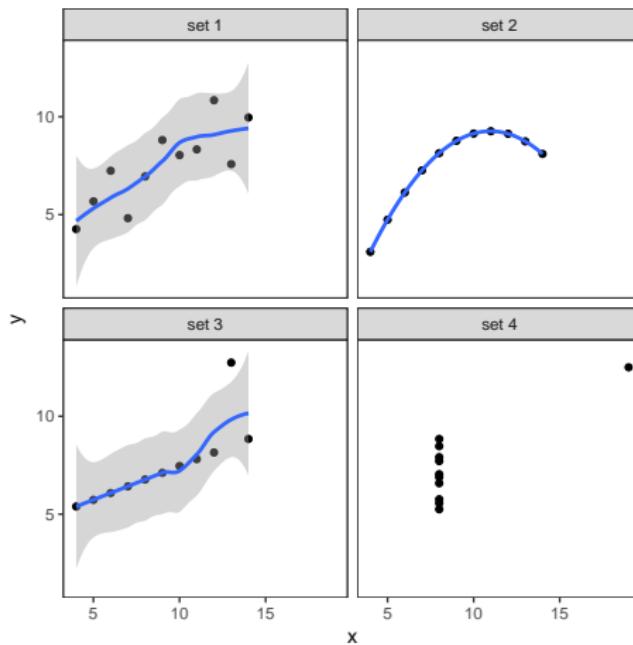
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



```
ggo = ggplot(anscombe, aes(x=x, y=y))
ggo = ggo + facet_wrap(~ set)
ggo = ggo + geom_point()

ggo = ggo + stat_smooth(method=loess, method.args=list(deg=1))
```

GSE75386 scatterplot matrix (a.k.a. pairs plot)

Data
Visualization

xkcd

History

Principles

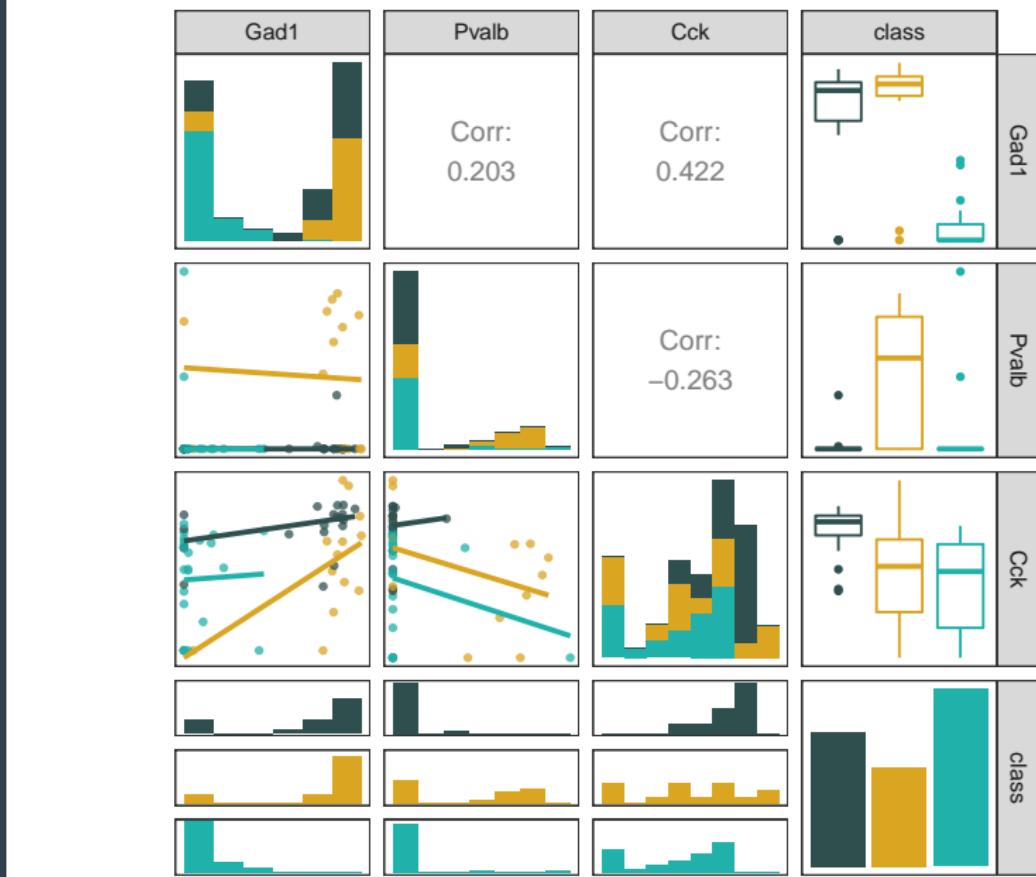
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



Clustered heatmap

Data
Visualization

xkcd

History

Principles

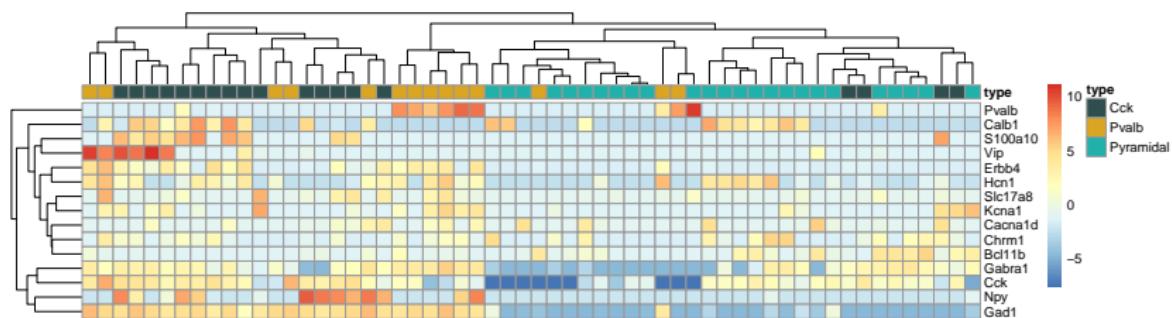
Grammar of
graphics

Minard

Composite
graphics

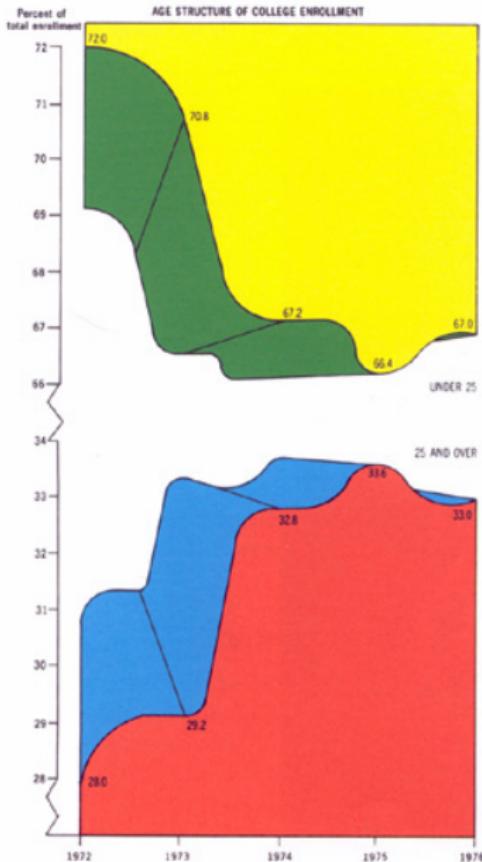
Rogue's
Gallery

References



```
pheatmap(  
    heatmapData,  
    annotation_col = data.frame(  
        row.names = colnames(heatmapData),  
        type = simpleType[colnames(heatmapData)]  
    ),  
    annotation_colors = list(type=c(  
        Cck = 'darkslategray',  
        Pvalb = 'goldenrod',  
        Pyramidal = 'lightseagreen'  
    )),  
    cluster_method = 'mcquitty',  
    show_colnames = FALSE  
)
```

Worst graph ever?



Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

Worst graph ever?

Data
Visualization

xkcd

History

Principles

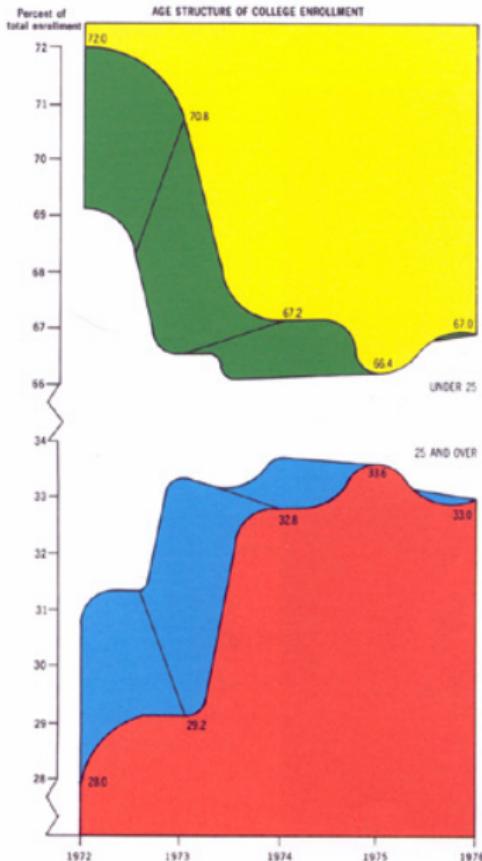
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



Tufte:

“Graphical excellence . . . gives the viewer

- ▶ the greatest number of ideas
- ▶ in the shortest time
- ▶ with the least ink
- ▶ in the smallest space.”

When in doubt, remove ink/pixels which do not directly map data

This one's not so good either...

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

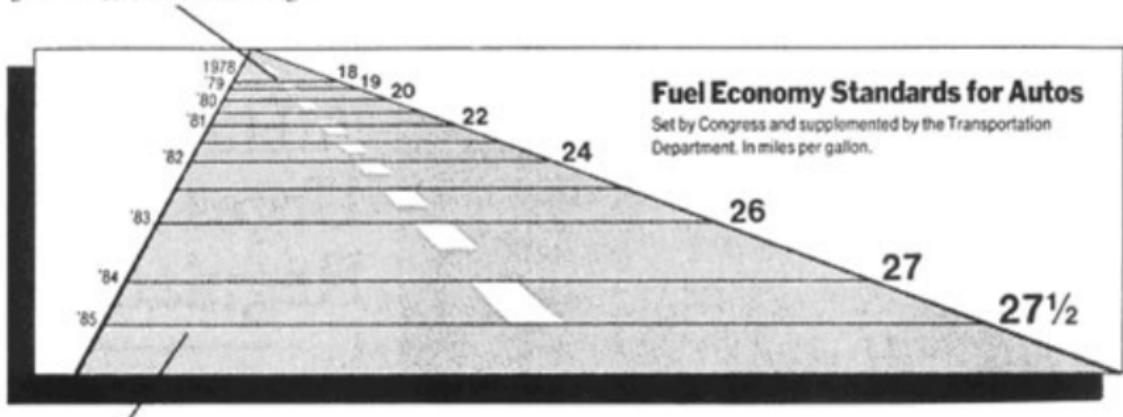
Minard

Composite
graphics

Rogue's
Gallery

References

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Moire effects

Data
Visualization

xkcd

History

Principles

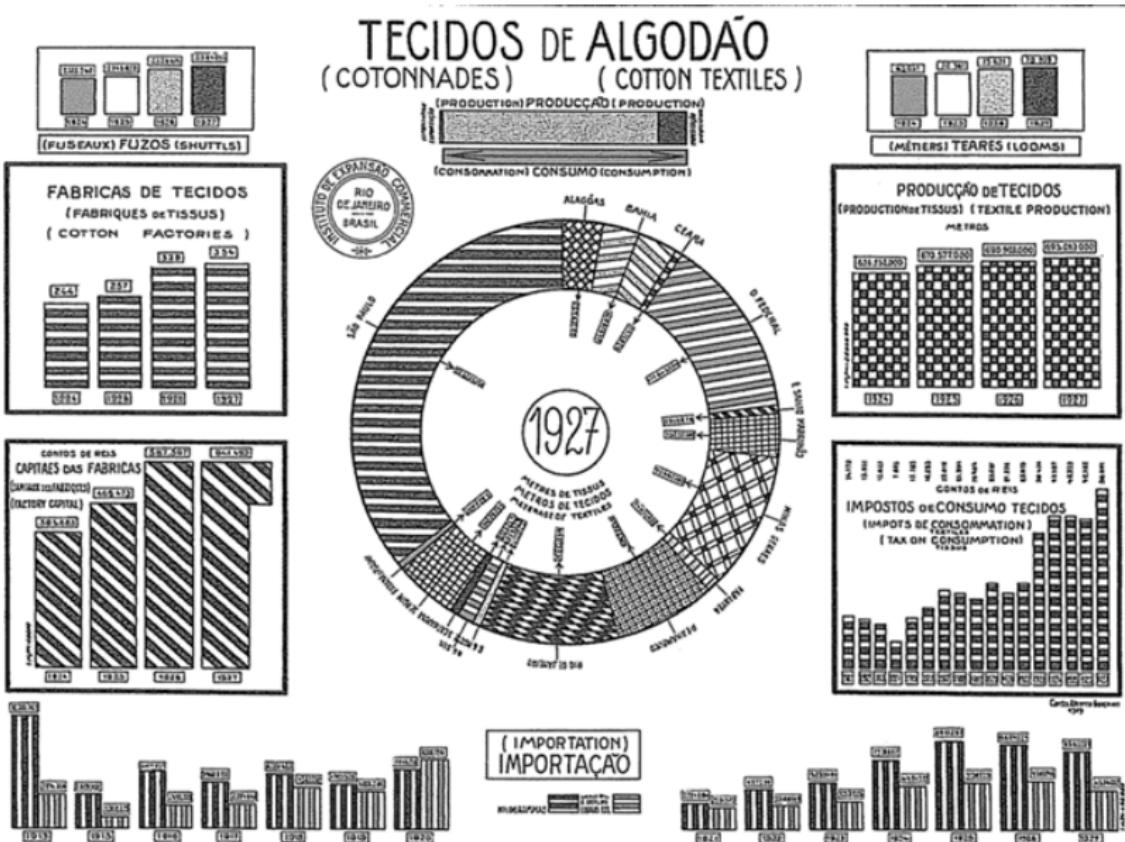
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



Moire effects

Data
Visualization

xkcd

History

Principles

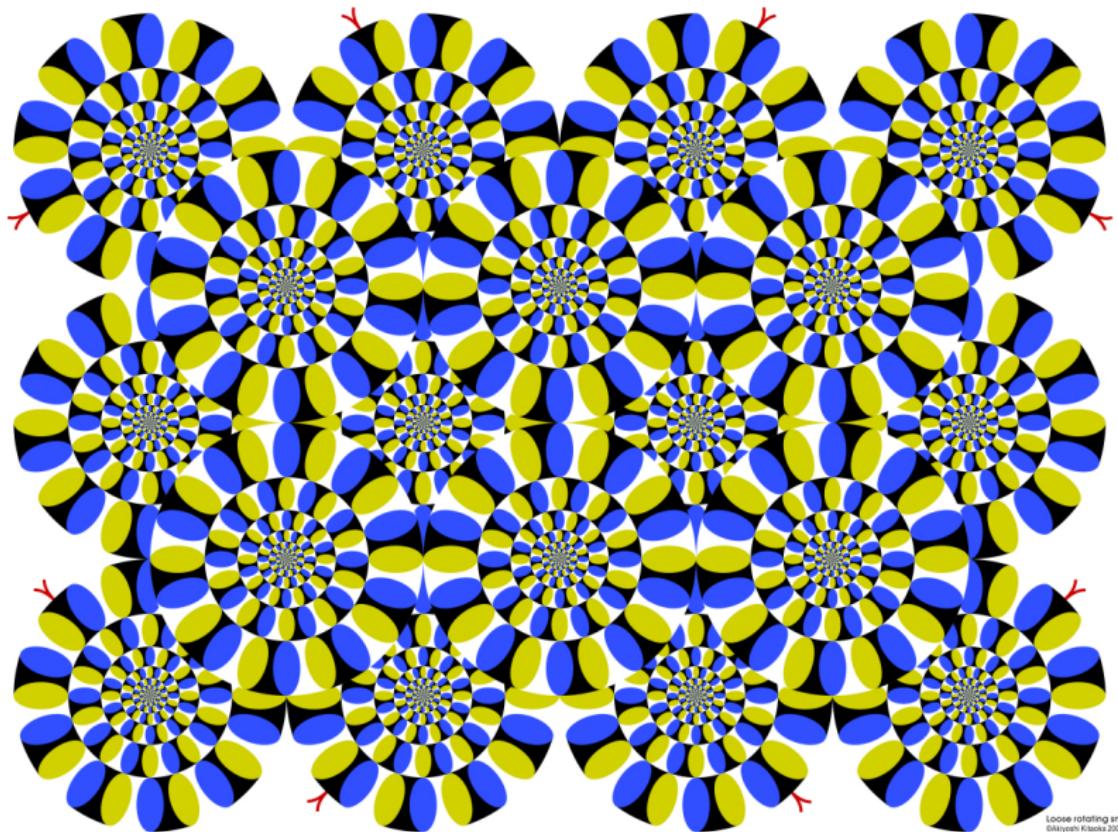
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References



Loose rotating snakes
©Kiyoshi Kitao 2009

References |

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

References

- Anscombe, Francis J. 1973. Graphs in statistical analysis. *The American Statistician*, **27**(1), 17–21.
- Friendly, Michael. 2008. A brief history of data visualization. *Handbook of Data Visualization*, 15–56.
- Hegarty, Mary. 2011. The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, **3**(3), 446–474.
- Mackinlay, Jock. 1986. Automating the design of graphical presentations of relational information. *ACM Transactions On Graphics (Tog)*, **5**(2), 110–141.
- Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.
- Wickham, Hadley. 2010. A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, **19**(1), 3–28.