

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

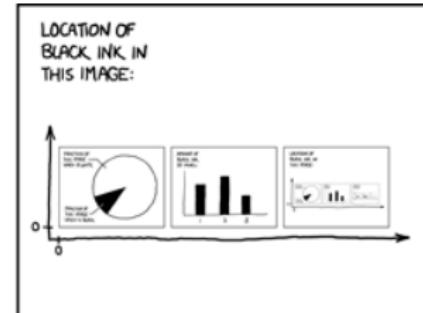
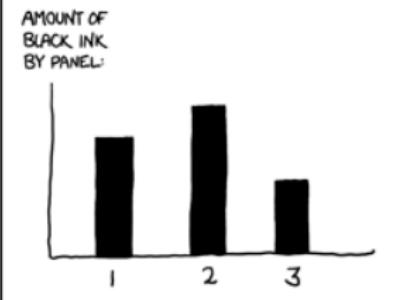
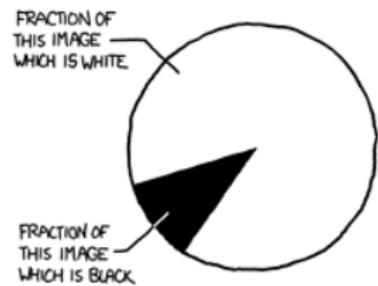
Composite
graphics

Rogue's
Gallery

Data Visualization

Dennis Wylie, UT Bioinformatics Consulting Group

April 11, 2017



Types of visualization

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

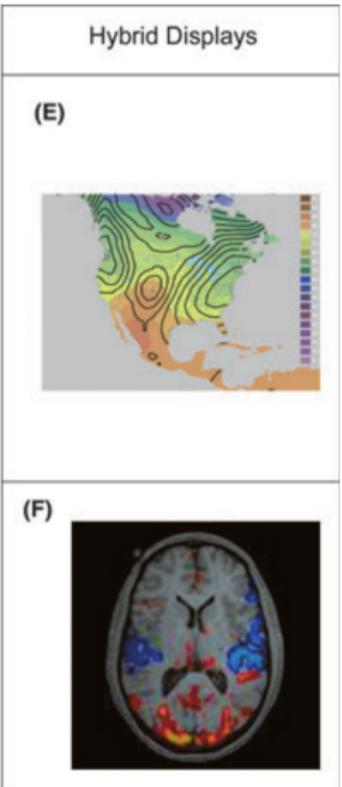
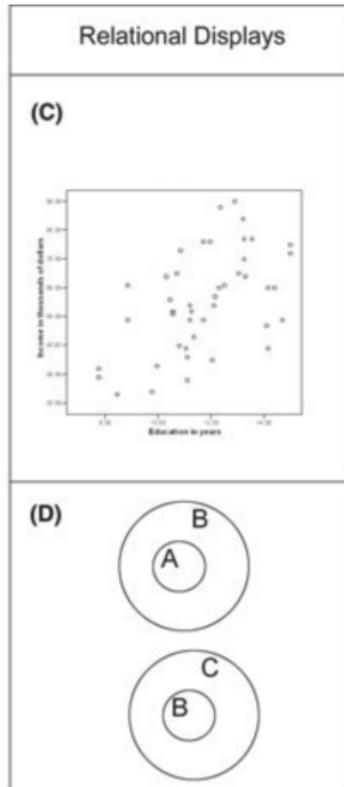
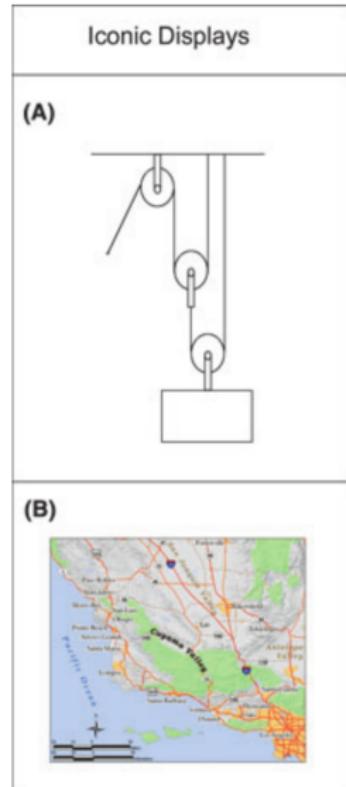


Fig. 1. Examples of iconic, relational, and hybrid displays.

History of data visualization

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

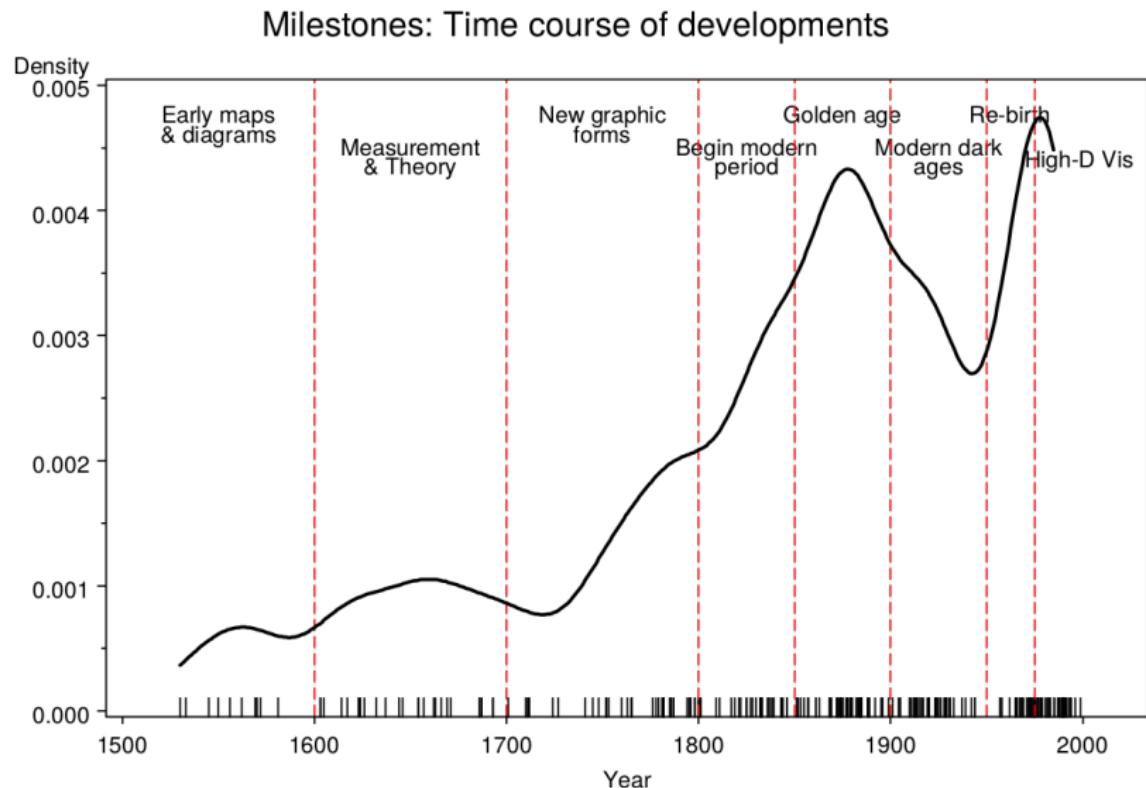


Figure 1: The time distribution of events considered milestones in the history of data visualization, shown by a rug plot and density estimate.

Playfair

Data
Visualization

xkcd

History

Principles

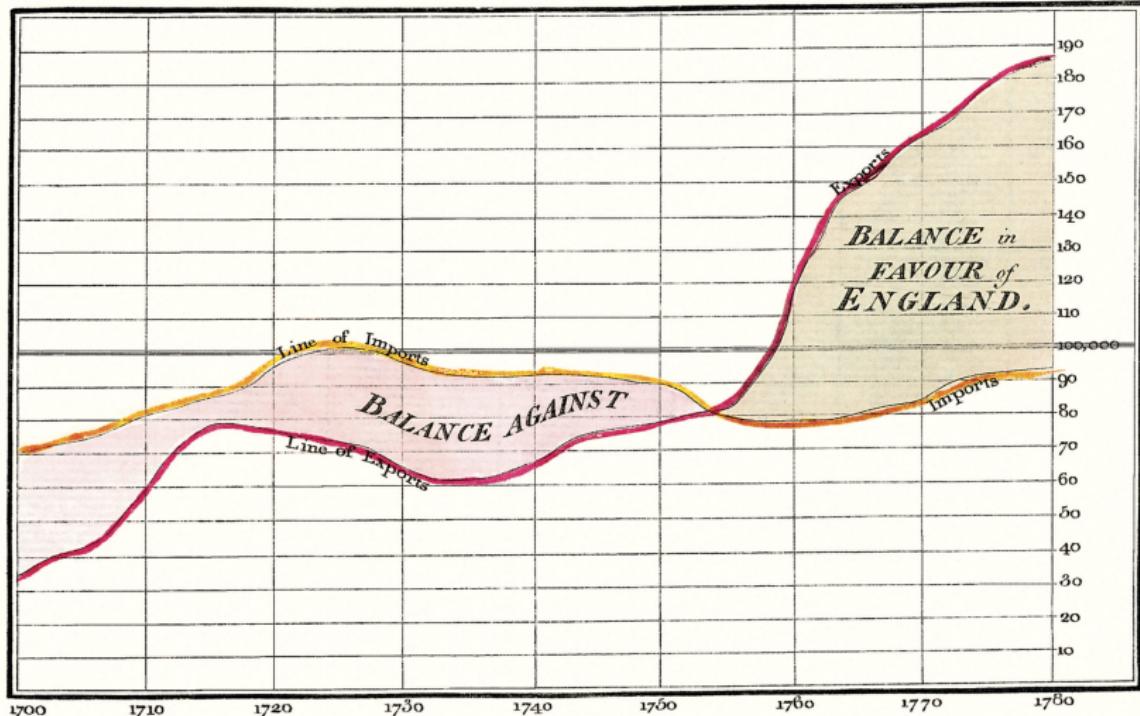
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

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

No. 62, Strand, London.

Minard

Data
Visualization

xkcd

History

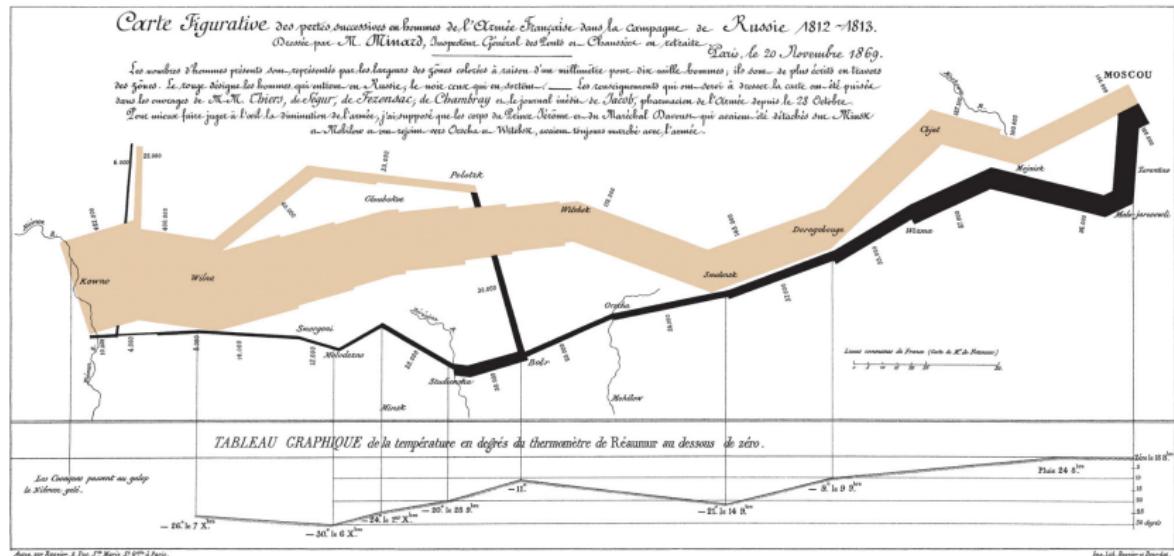
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



Nightingale

Data
Visualization

xkcd

History

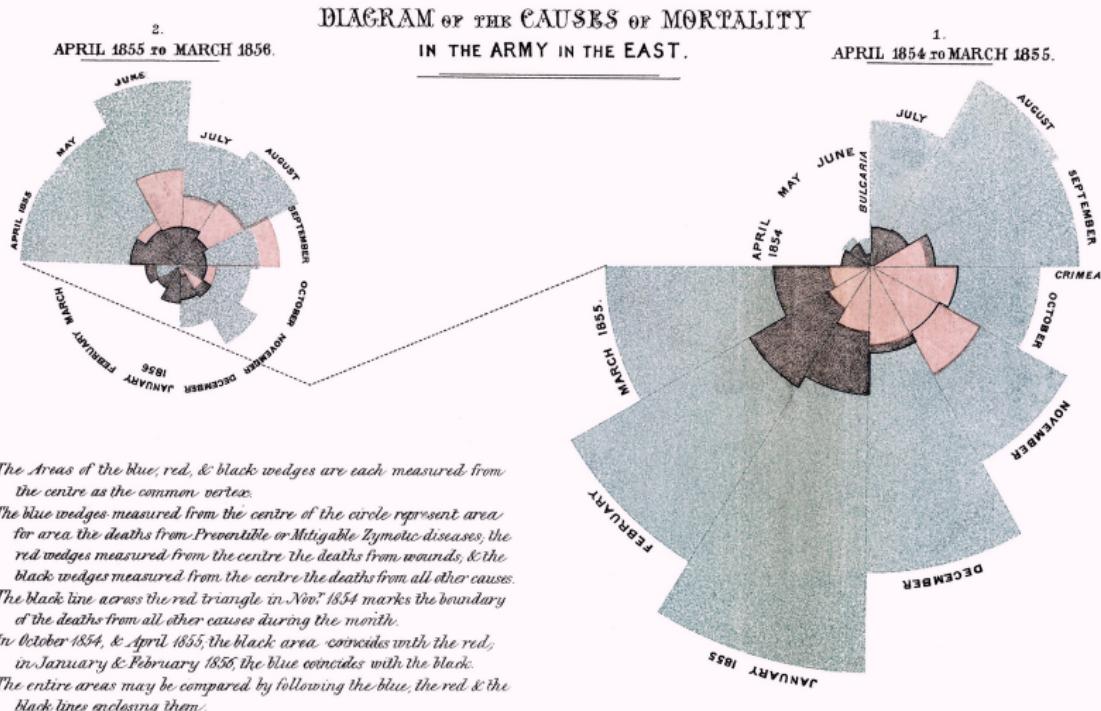
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



Why visualize?

Data
Visualization

- ▶ Summarize data

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

Why visualize?

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

- ▶ Summarize data
- ▶ Illustrate relationships

Why visualize?

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

- ▶ 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

- ▶ 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

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

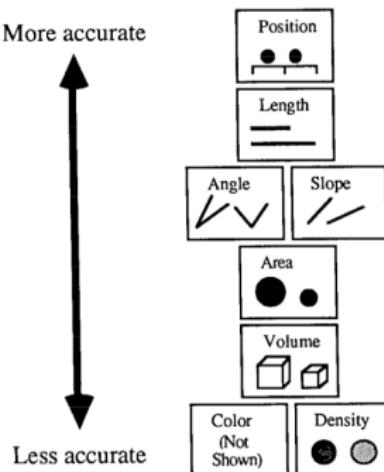


Fig. 14. Accuracy ranking of quantitative perceptual tasks. Higher tasks are accomplished more accurately than lower tasks. Cleveland and McGill empirically verified the basic properties of this ranking.

Natural example: frequency mapped to angle

Data
Visualization

xkcd

History

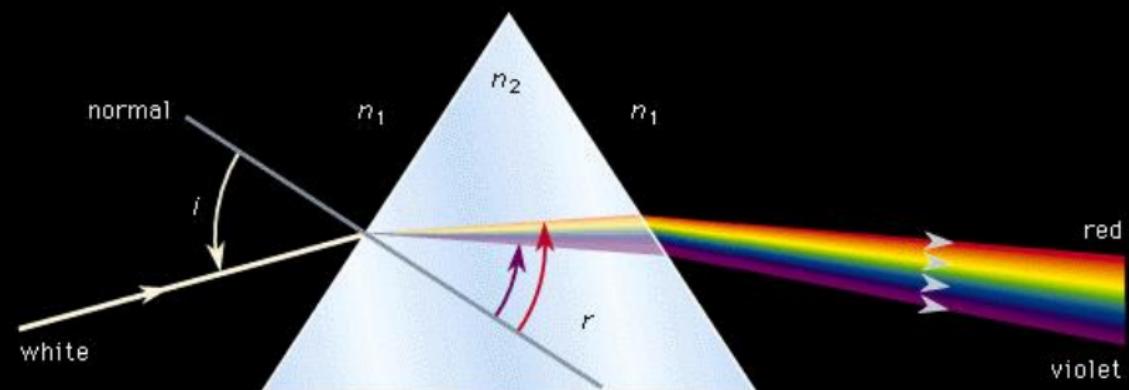
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



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.

©1996 Encyclopaedia Britannica, Inc.

Experimental example: size mapped to position

Data
Visualization

xkcd

History

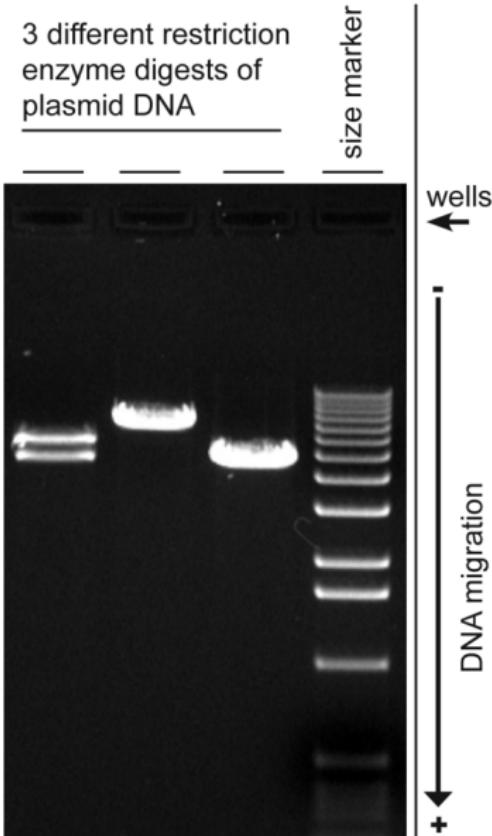
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



- ▶ 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

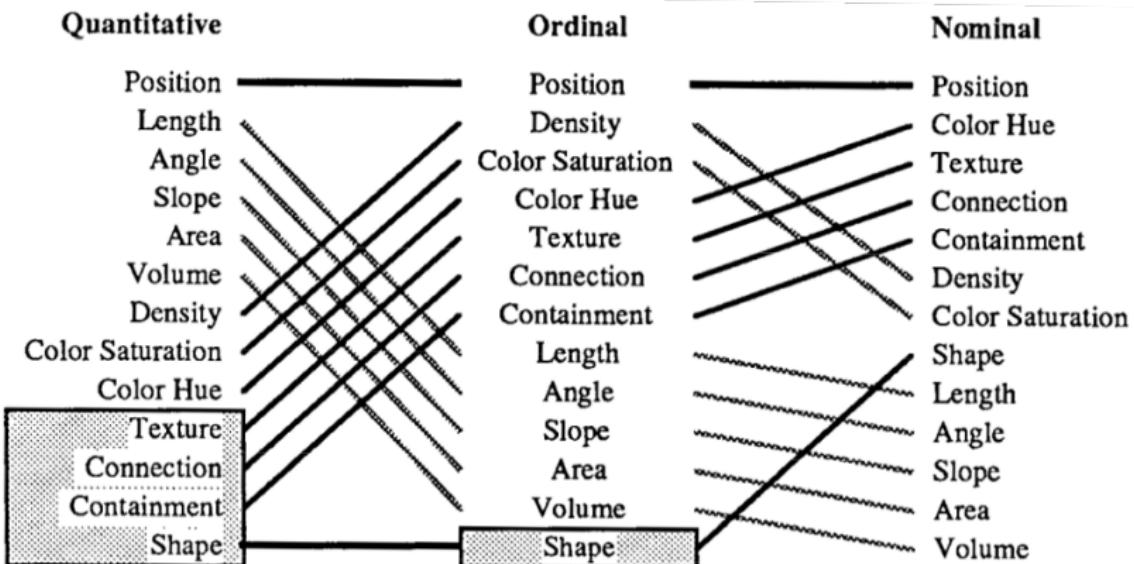


Fig. 15. Ranking of perceptual tasks. The tasks shown in the gray boxes are not relevant to these types of data.

"Lie factor" from improper area encoding

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

	1964	1975	1990
	27 %	16.0 %	12.0 %



Neocortical neuron subtypes and GSE75386

Data
Visualization

xkcd

History

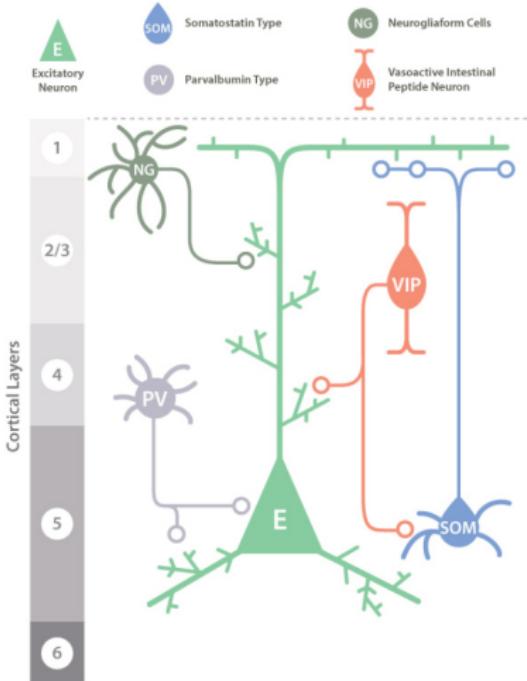
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



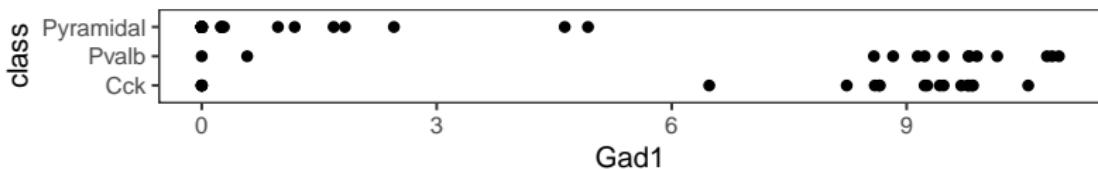
© Knowing Neurons <http://knowingneurons.com>

GSE75386 Single cell
RNAseq of elec-
trophysiologically
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

GSE75386 stripchart example

Data
Visualization

xkcd

History

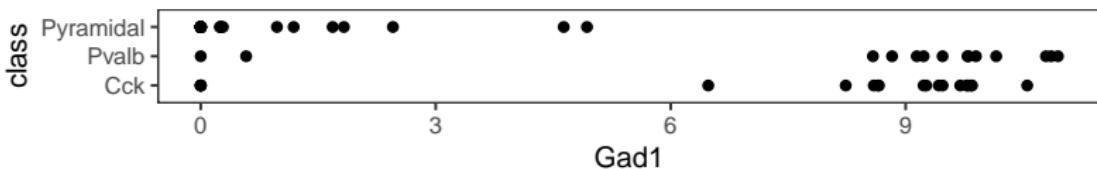
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



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

GSE75386 stripchart example

Data
Visualization

xkcd

History

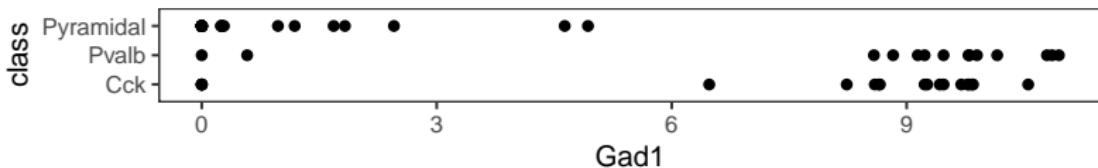
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



```
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



```
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



```
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

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

layer = data +
statistical transformation +
aesthetic +
geometric object

GSE75386 boxplot + stripchart

Data
Visualization

xkcd

History

Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



```
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



```
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

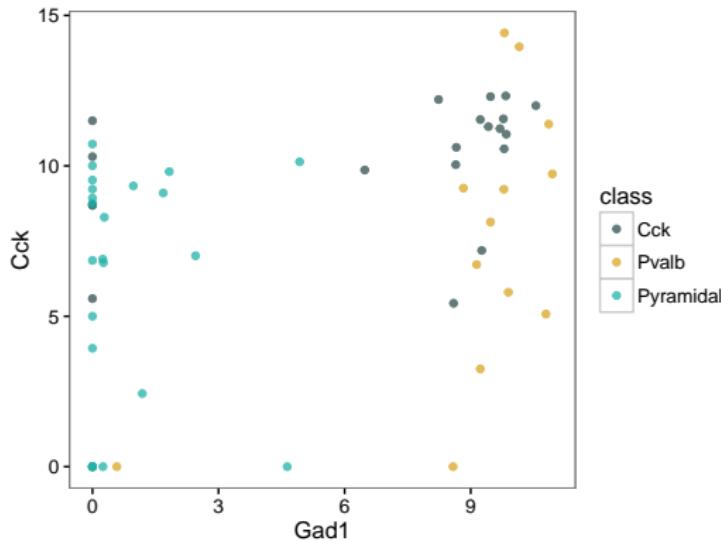
Data Visualization

History

Principles

Grammar of graphics

Minard



```
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

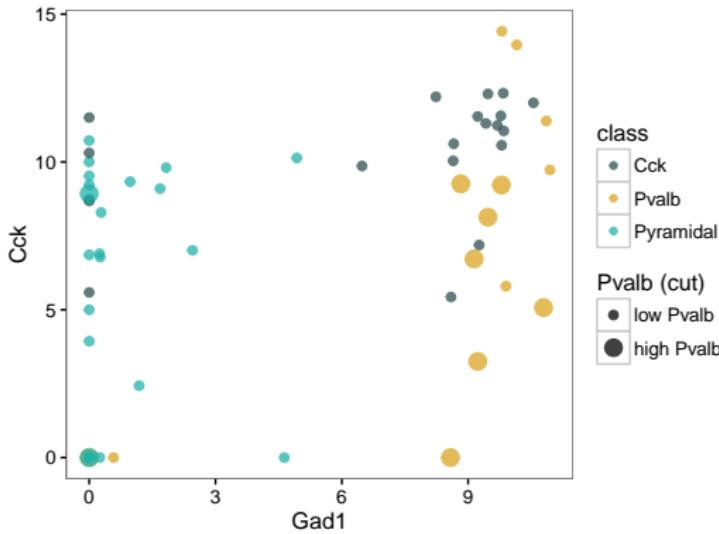
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



```
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

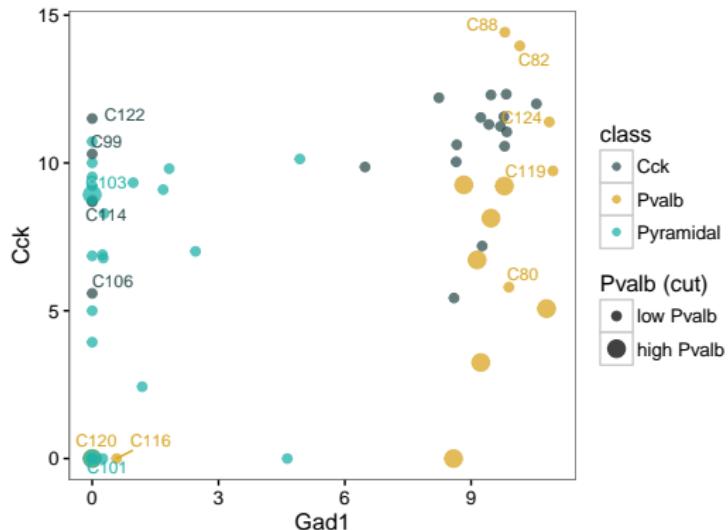
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



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

The best graph ever?

Data
Visualization

xkcd

History

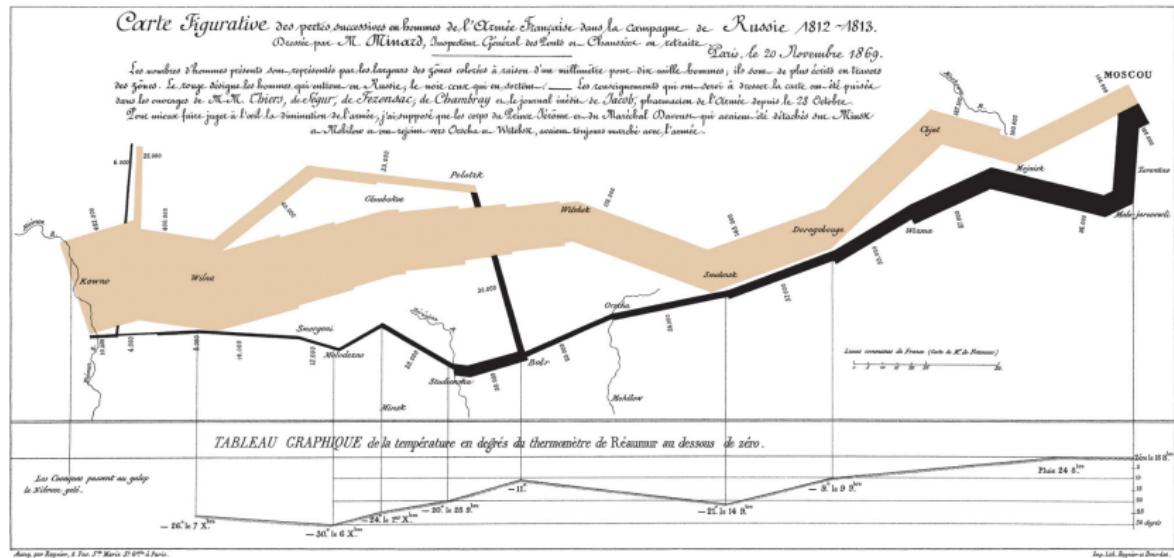
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



ggplotting Minard

Data
Visualization

xkcd

History

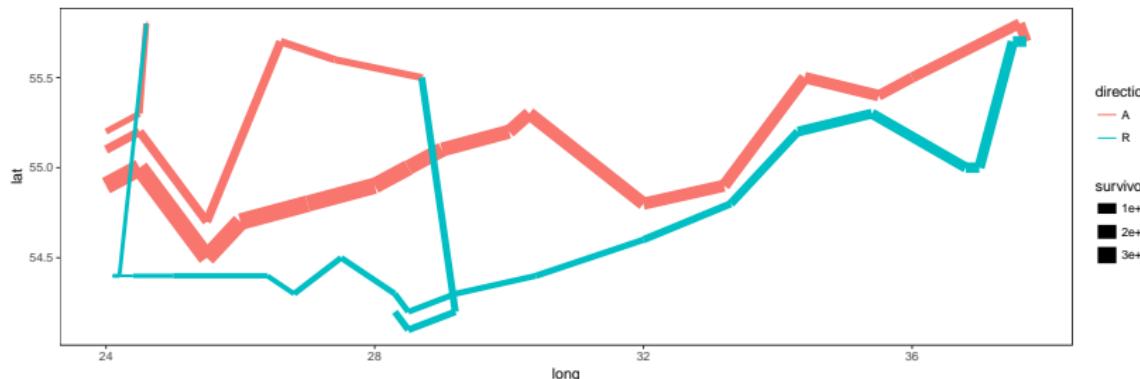
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



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

ggplotting Minard

Data
Visualization

xkcd

History

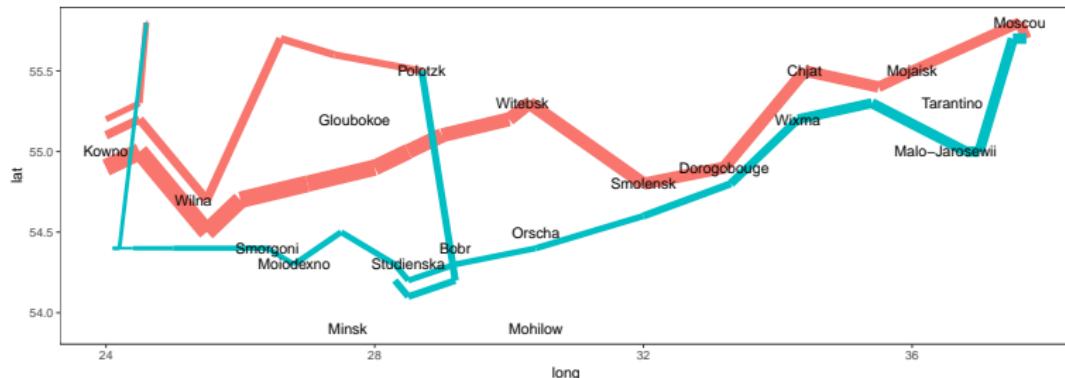
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



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

ggplotting Minard

Data
Visualization

xkcd

History

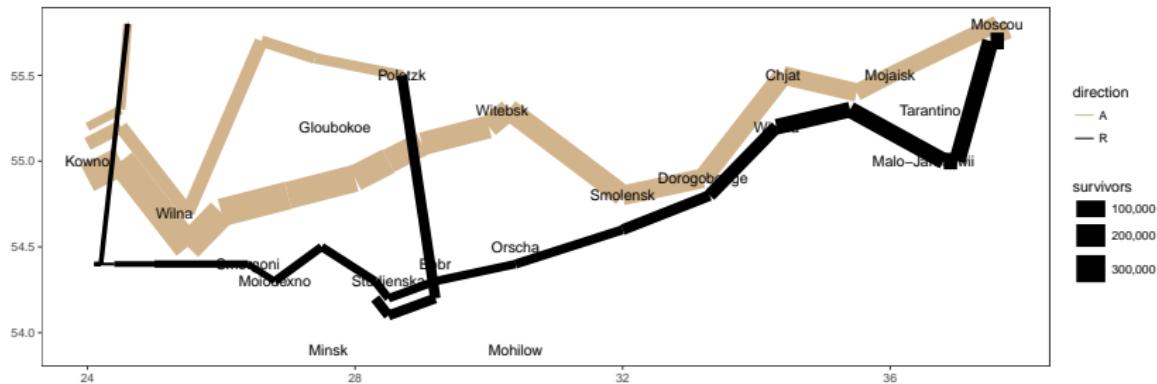
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



```
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

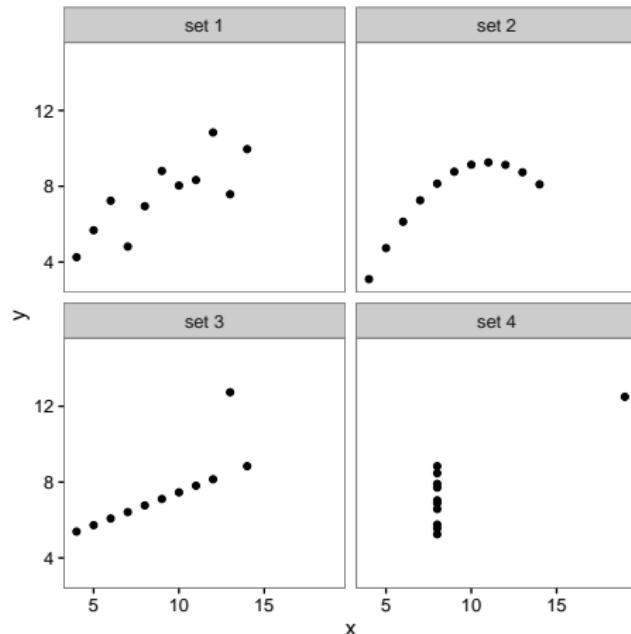
Principles

Grammar of
graphics

Minard

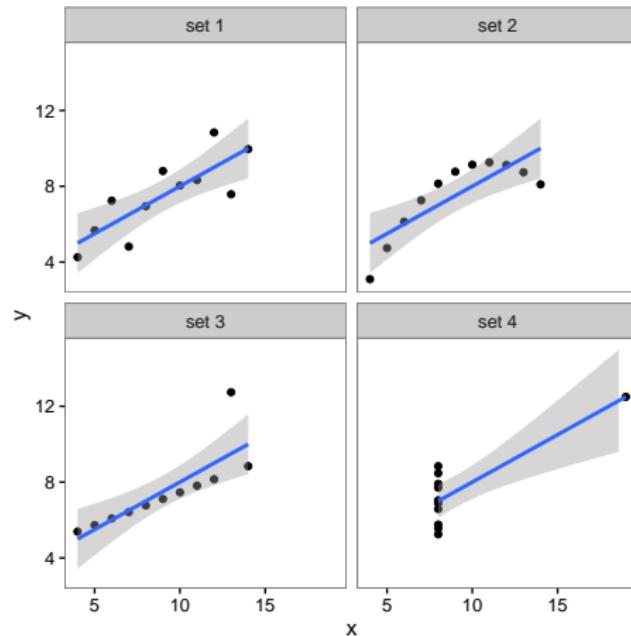
Composite
graphics

Rogue's
Gallery



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

... with smoothed linear fit



```
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

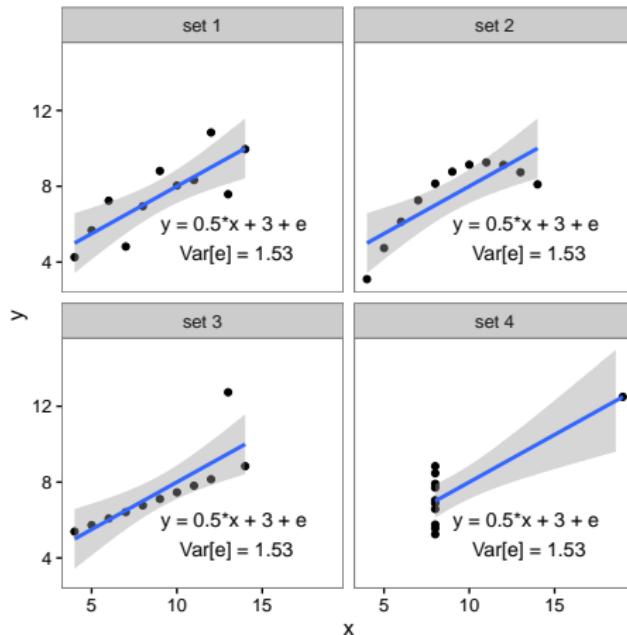
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



```
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

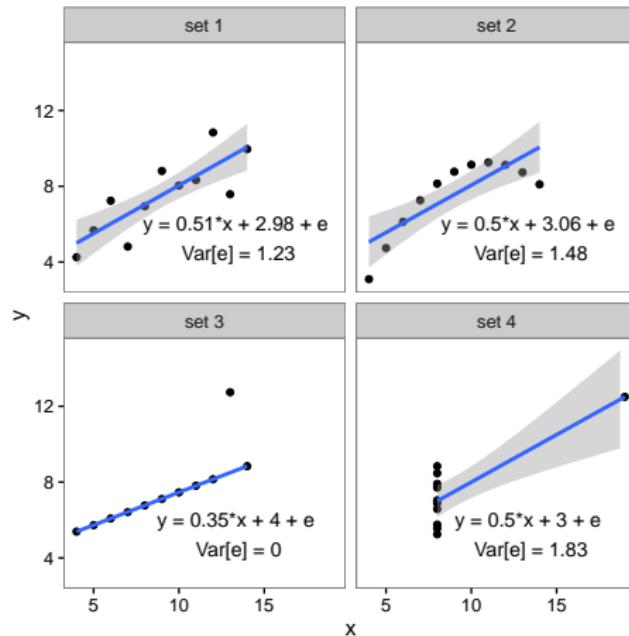
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



```
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

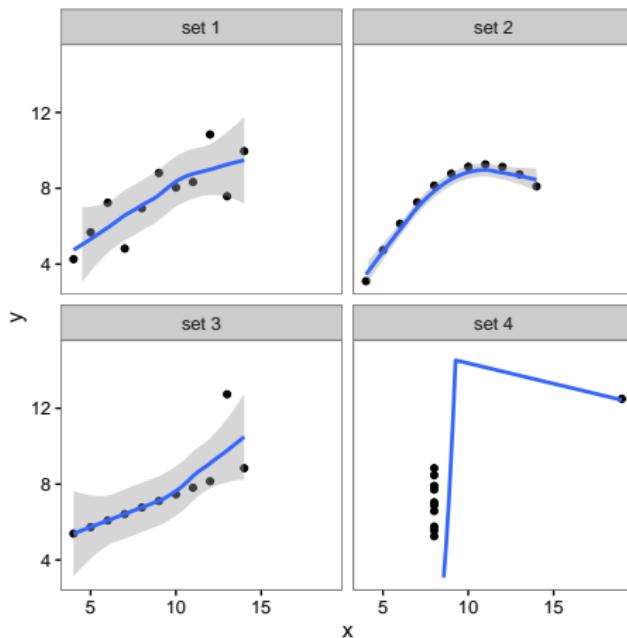
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



```
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

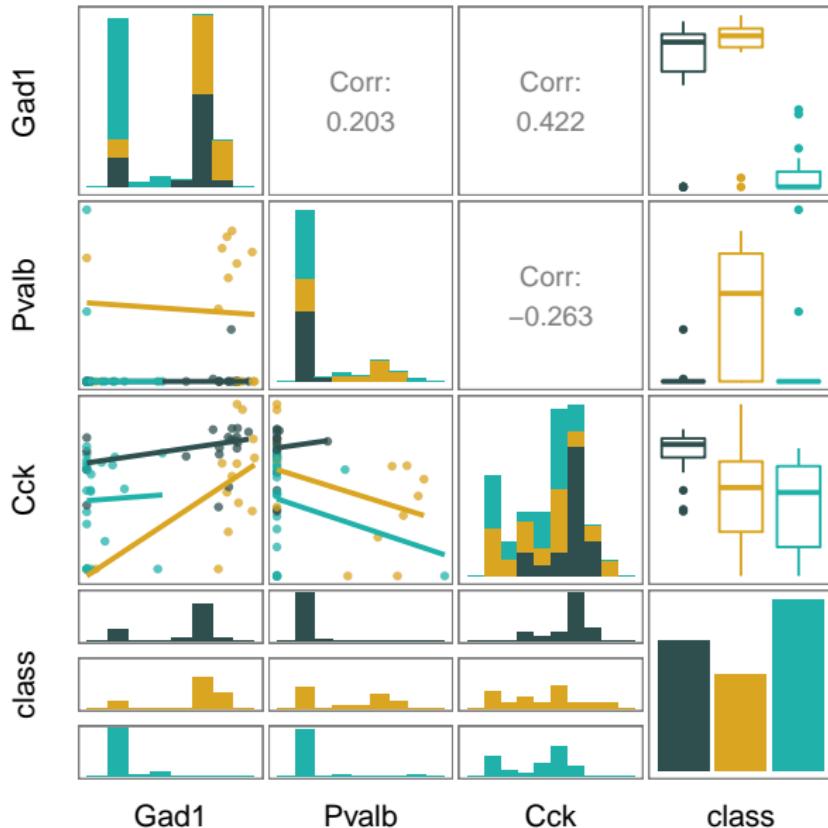
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



Clustered heatmap

Data
Visualization

xkcd

History

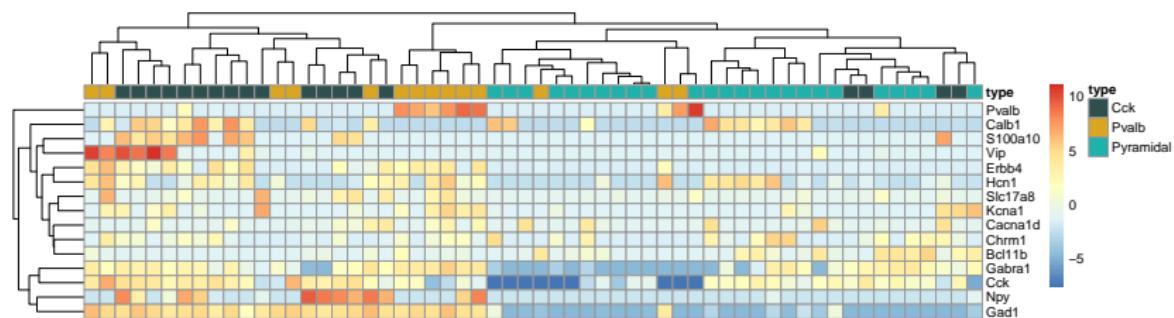
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



```
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

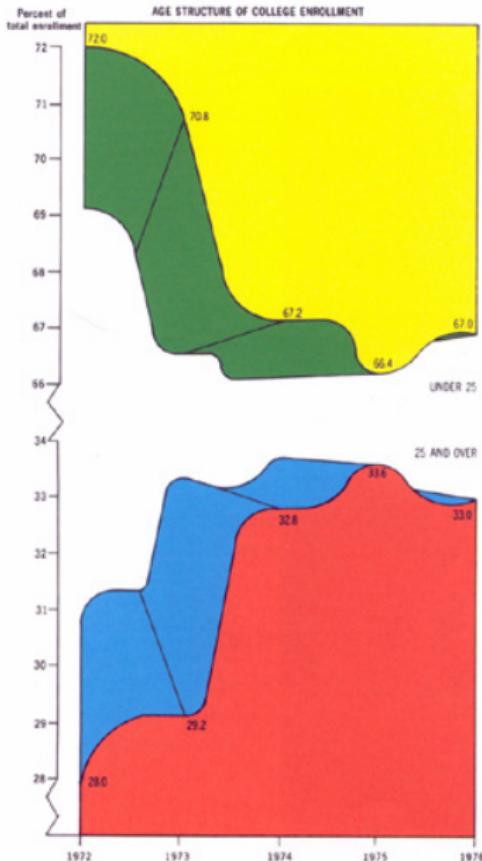
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



Worst graph ever?

Data
Visualization

xkcd

History

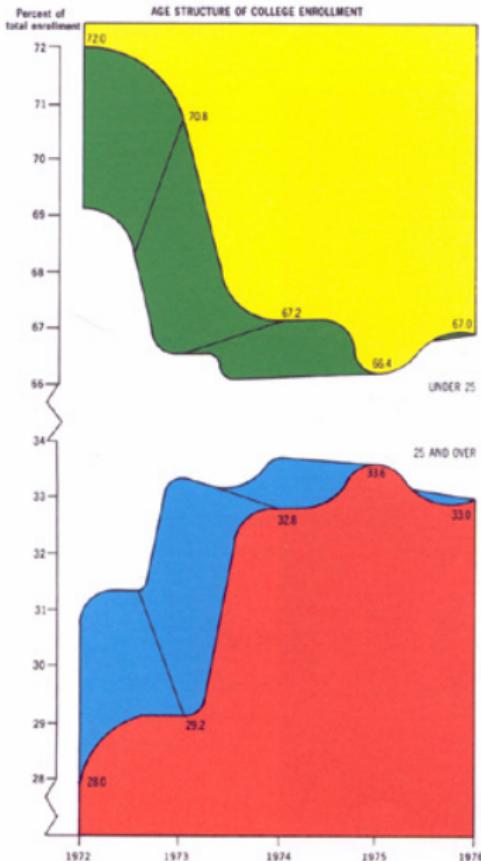
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



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

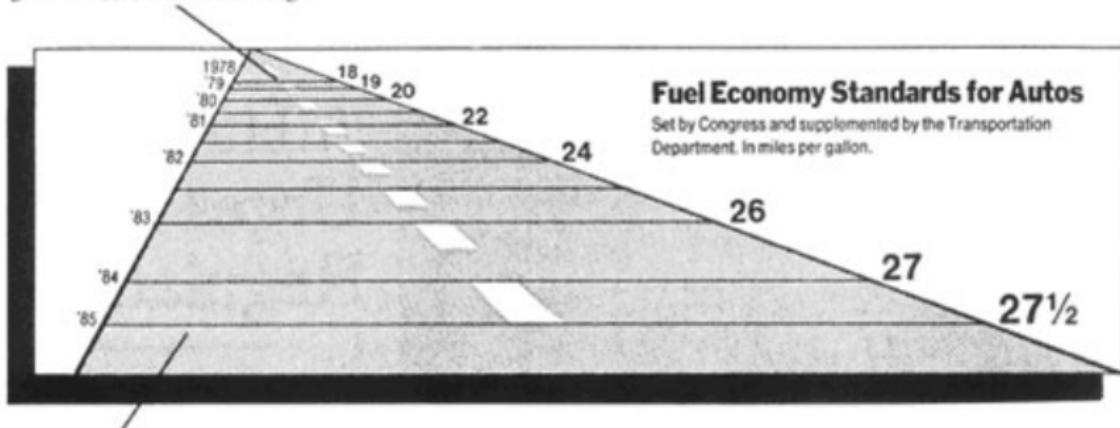
Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery

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

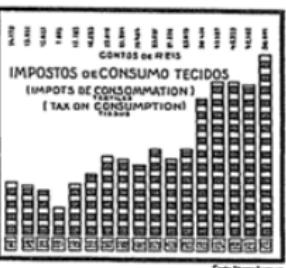
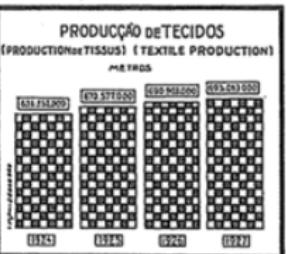
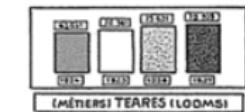
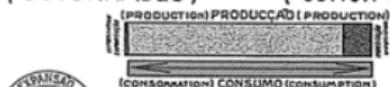
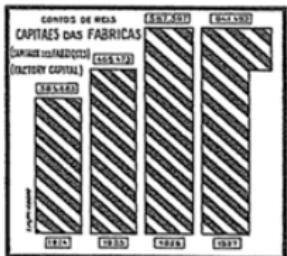
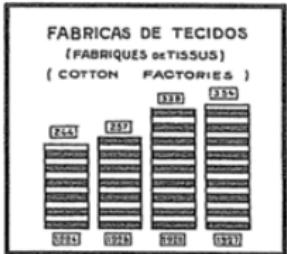
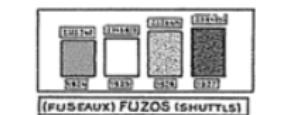
xkcd

History

Principles

Minard

Rogue's
Gallery



Moire effects

Data
Visualization

xkcd

History

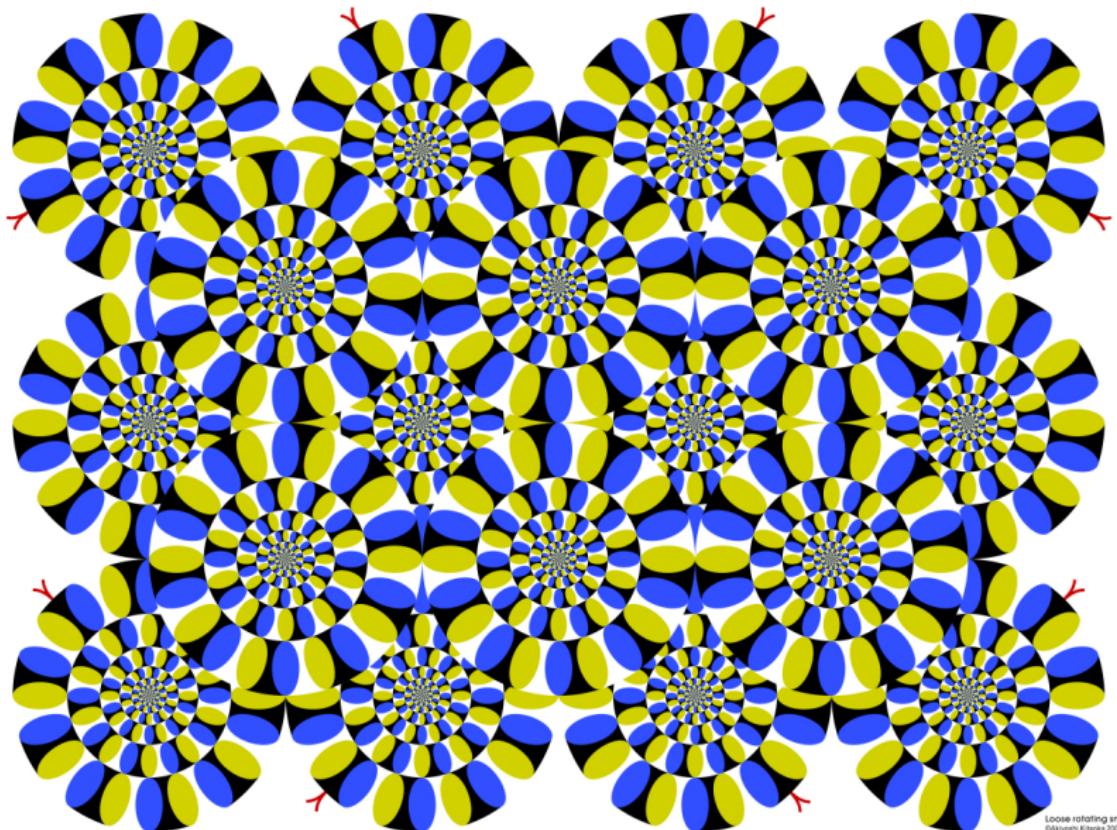
Principles

Grammar of
graphics

Minard

Composite
graphics

Rogue's
Gallery



Loose rotating snakes
©Mitsuru Kitajima 2009