

# Cleveland's Principles

---

## “Clear vision”

- Data should stand out—nothing superfluous
- Use visual elements to highlight the data
- Declutter
- Include reference lines

Source: Cleveland, *The Elements of Graphing Data*, 1985

# Cleveland's Principles

“Clear understanding”

Conclusions should be in graphical form

Legends should inform (comprehensively!)

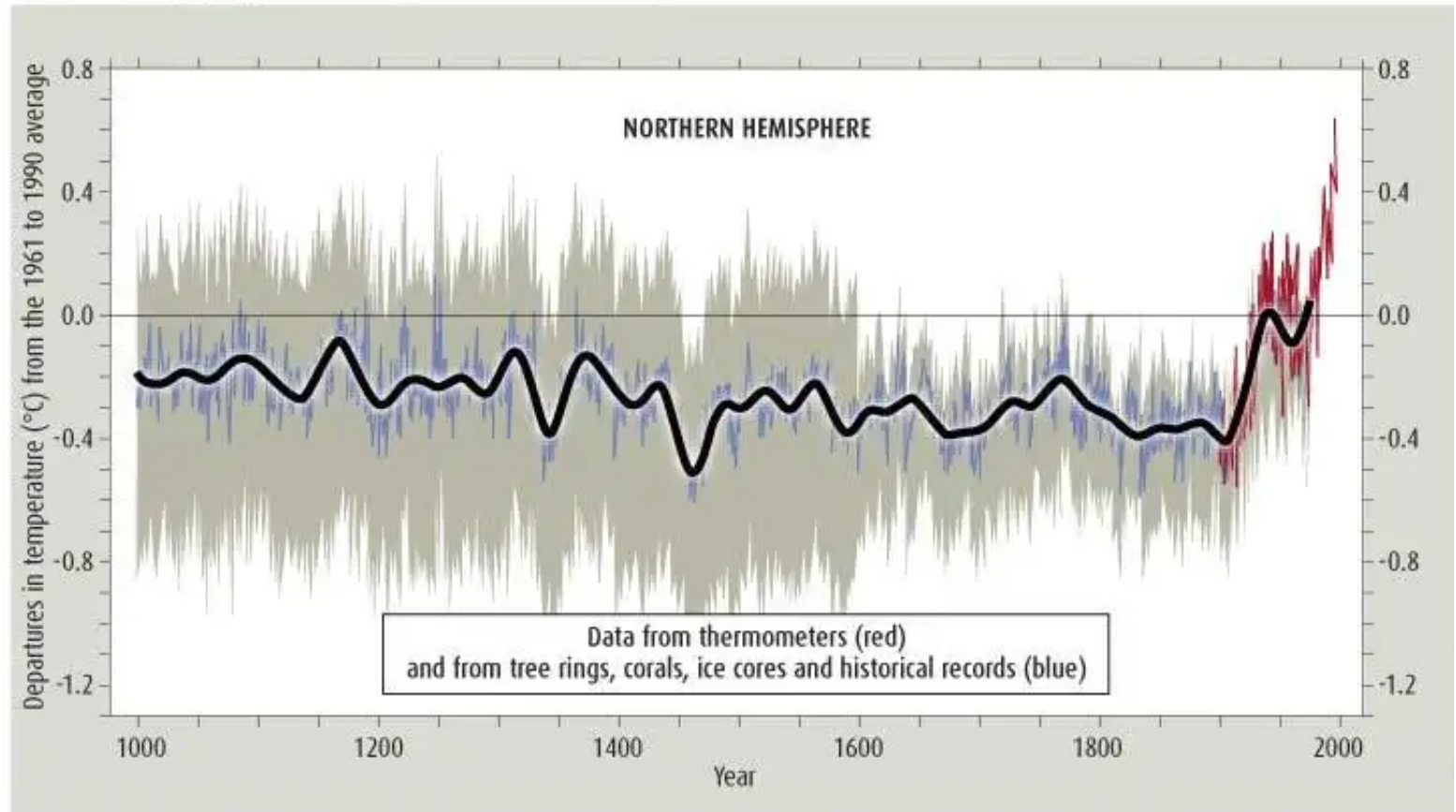
Scales

- consider limits, logarithmic, dual, etc.

# Albert Cairo, *The Truthful Art* [2016]

- Five qualities of “great visualizations” (p. 45)
  - truthful
  - Functional
  - Beautiful
  - Insightful
  - enlightening

# “Variations of the Earth’s surface temperature over the past 1000 years”



Source: <https://www.newscientist.com/article/dn11646-climate-myths-the-hockey-stick-graph-has-been-proven-wrong/>

# Discussion

*Figures on next 3 slides are from Introduction of Spiegelhalter's Art of Statistics*

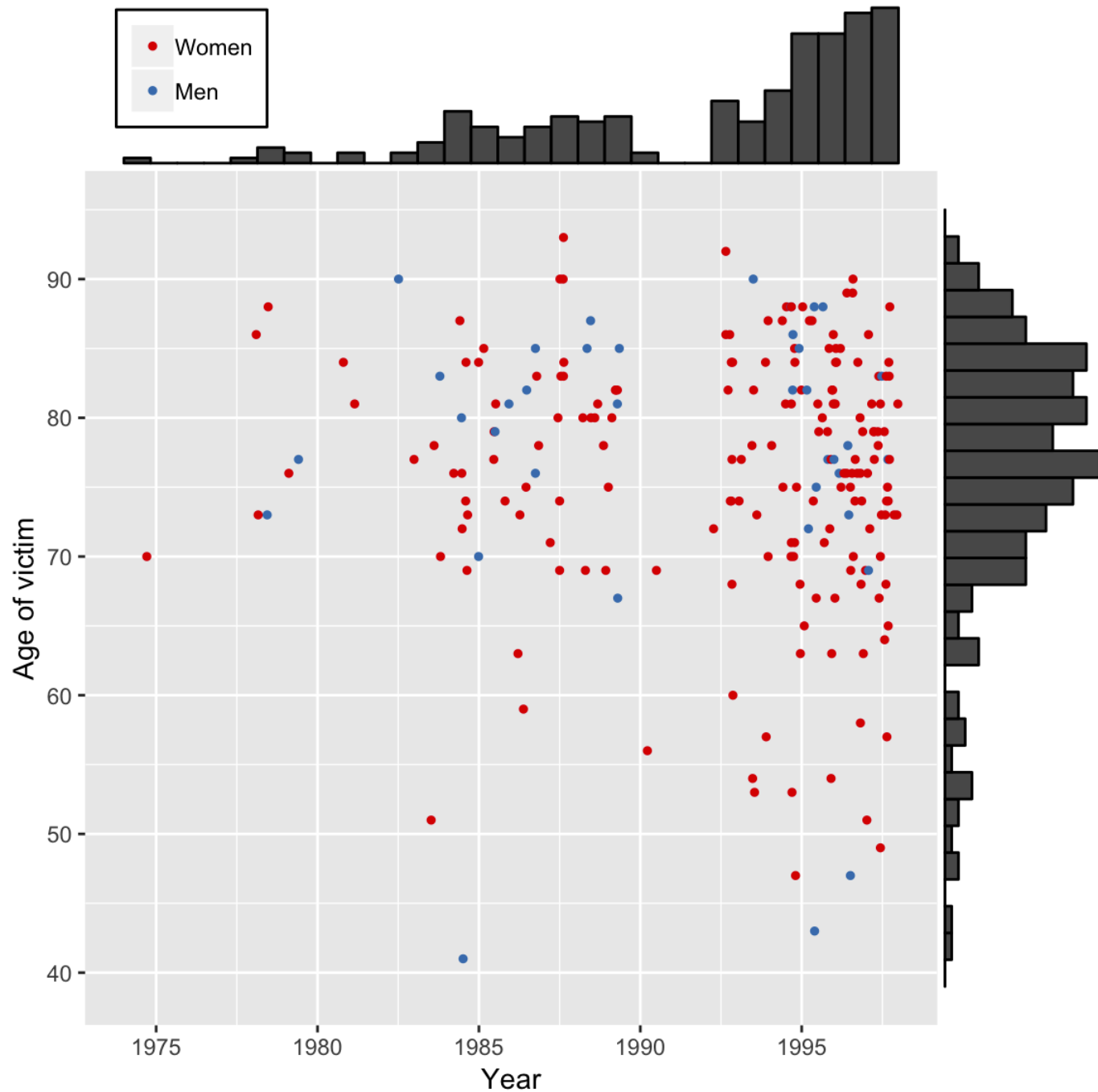
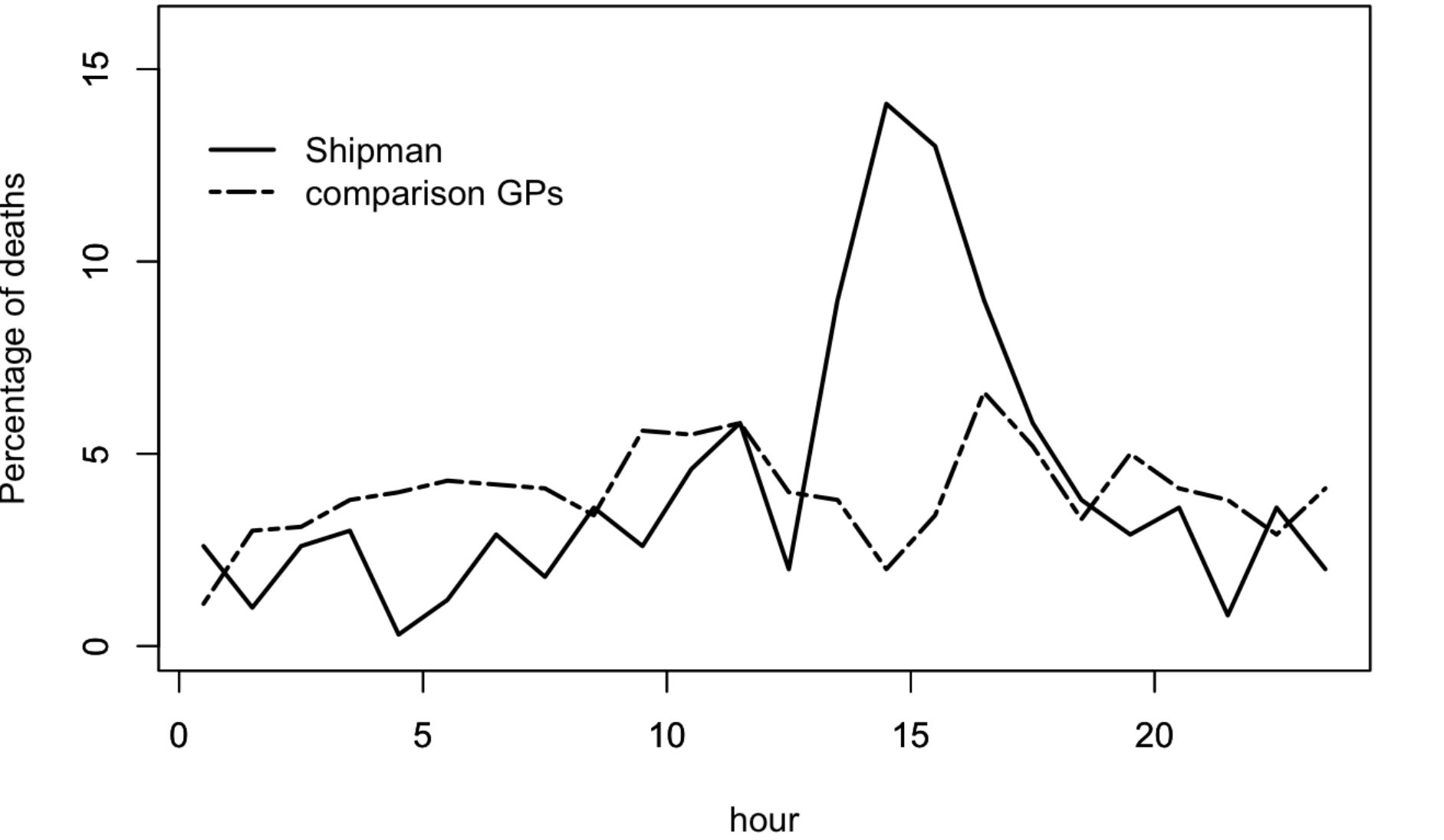


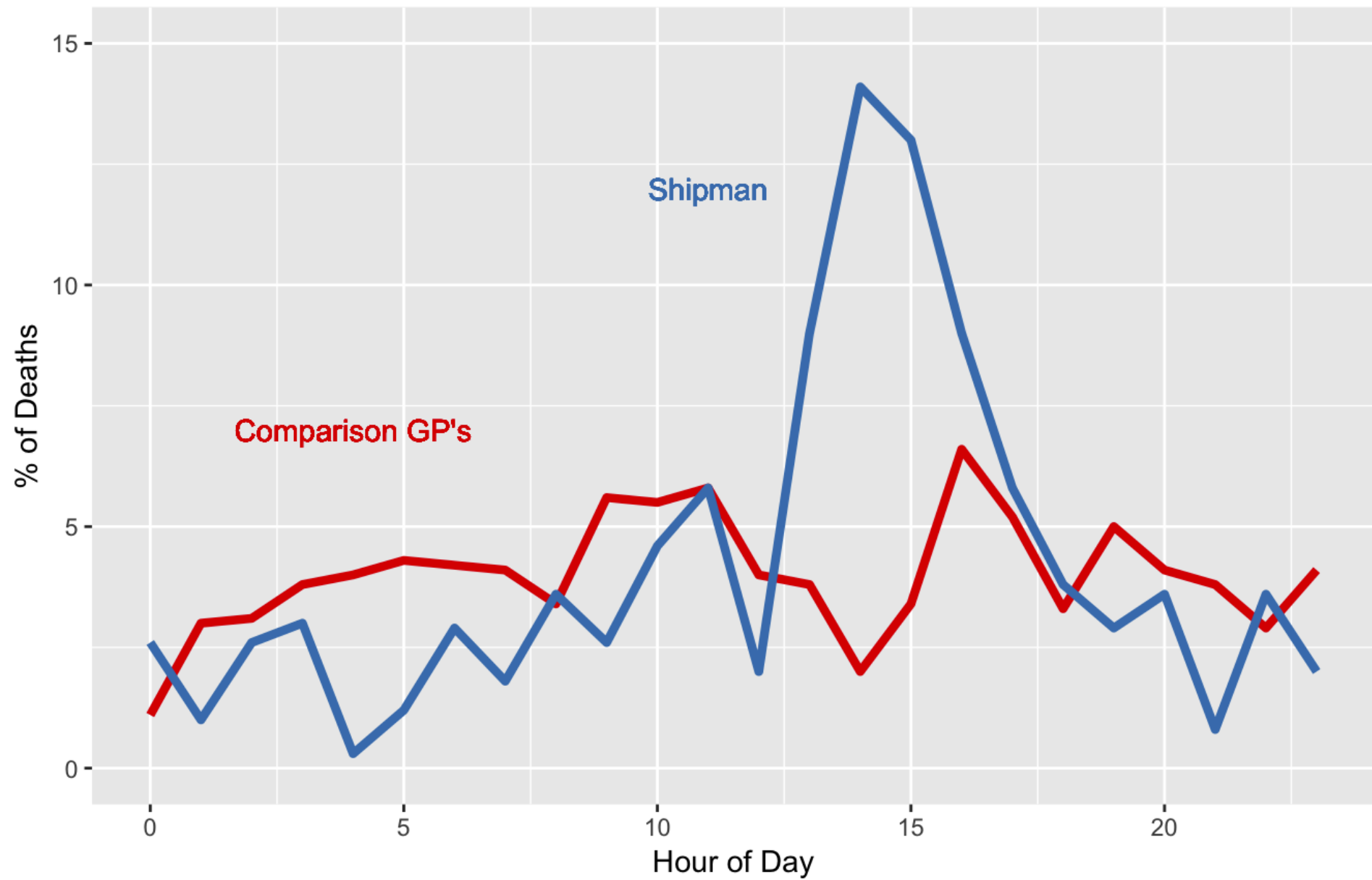
Figure 0.1 A scatter-plot showing the age and the year of death of Harold Shipman's 215 confirmed victims. Bar-charts have been added on the axes to reveal the pattern of ages and the pattern of years in which he committed murders.

Figure 0.2. The time at which Harold Shipman's patients died, compared to the times at which patients of other local general practitioners died. The pattern does not require sophisticated statistical analysis.



# Deaths by Hour of Day

From Shipman dataset





# (layered) grammar of graphics

- See [“A Layered Grammar of Graphics”](#)  
[Wickham 2010]
- Builds on “The Grammar of Graphics” by  
Wilkinson, Anand, and Grossman (2005)

# ggplot2

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

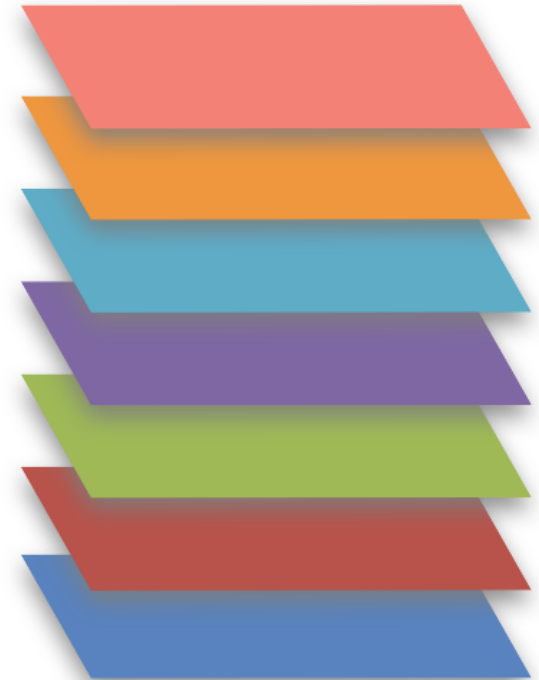
Statistics

Facets

Geometries

Aesthetics

Data



- Source: <https://towardsdatascience.com/murdering-a-legendary-data-story-what-can-we-learn-from-a-grammar-of-graphics-ad6ca42f5e30>

```
library(car); data(SLID)
```

# data

```
g <- ggplot(data=SLID)
```

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

Statistics

Facets

Geometries

Aesthetics

Data



# aesthetic

```
g <- ggplot(data=SLID,  
            aes(x=education, y=wages) )
```

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

Statistics

Facets

Geometries

Aesthetics

Data



# geometry

```
g <- ggplot(data=SLID,  
            aes(x=education, y=wages)) +  
  geom_point()
```

g

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

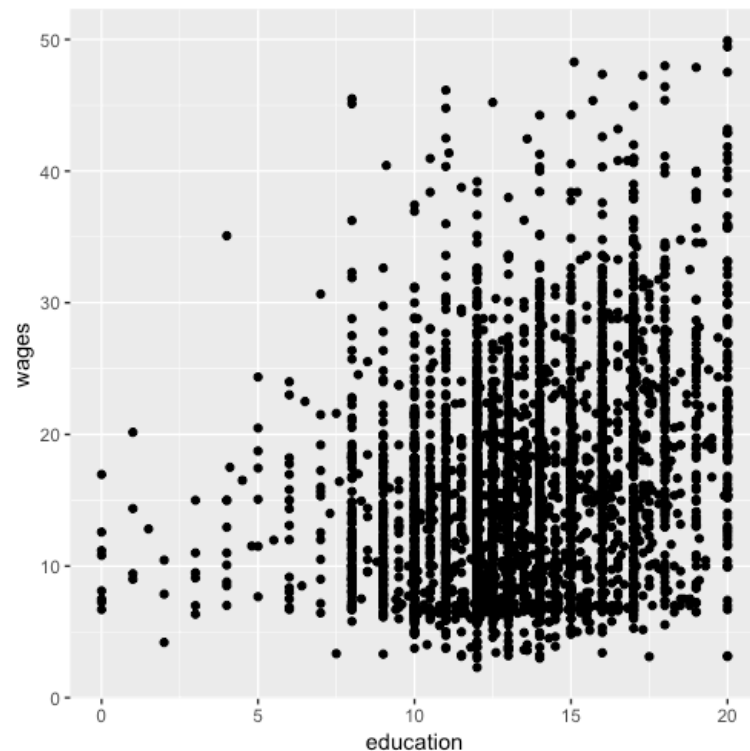
Statistics

Facets

Geometries

Aesthetics

Data



# geometry

```
g <- ggplot(data=SLID,  
            aes(x=education, y=wages)) +  
  geom_smooth()
```

g

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

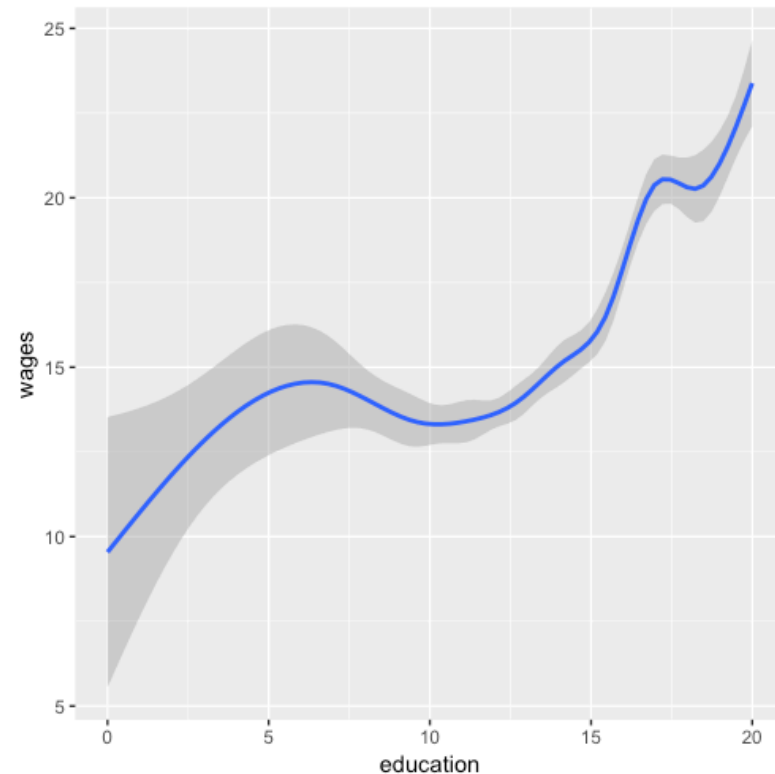
Statistics

Facets

Geometries

Aesthetics

Data



# geometry

```
g <- ggplot(data=SLID,  
            aes(x=education, y=wages)) +  
  geom_smooth()  
  
g <- g + geom_point()  
g
```

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

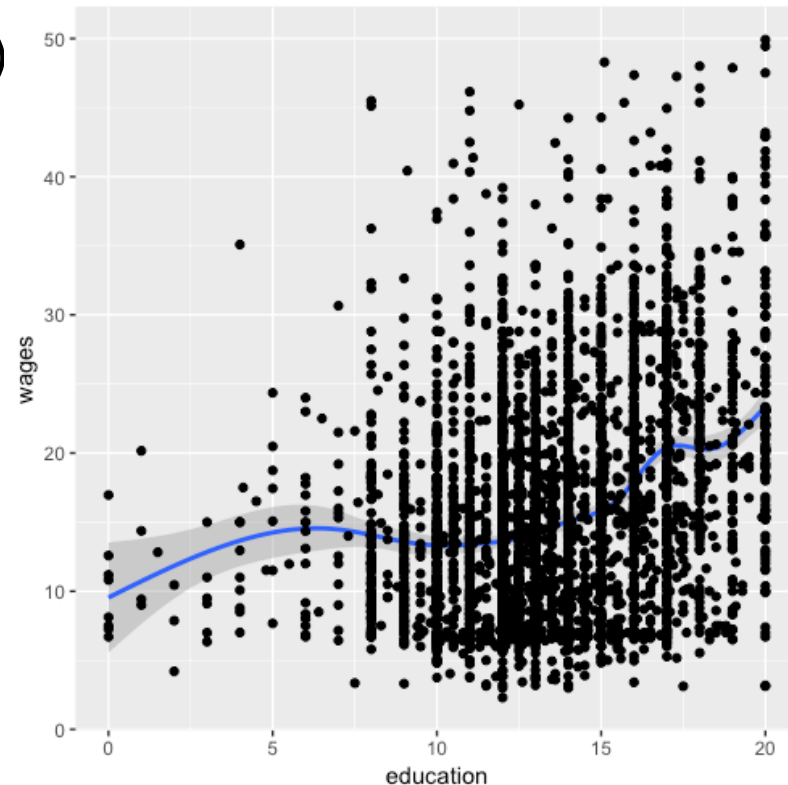
Statistics

Facets

Geometries

Aesthetics

Data



# aesthetic (again)

```
g <- ggplot(data=SLID)
g <- g + geom_point(
  aes(x=education, y=wages) )
g
```

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

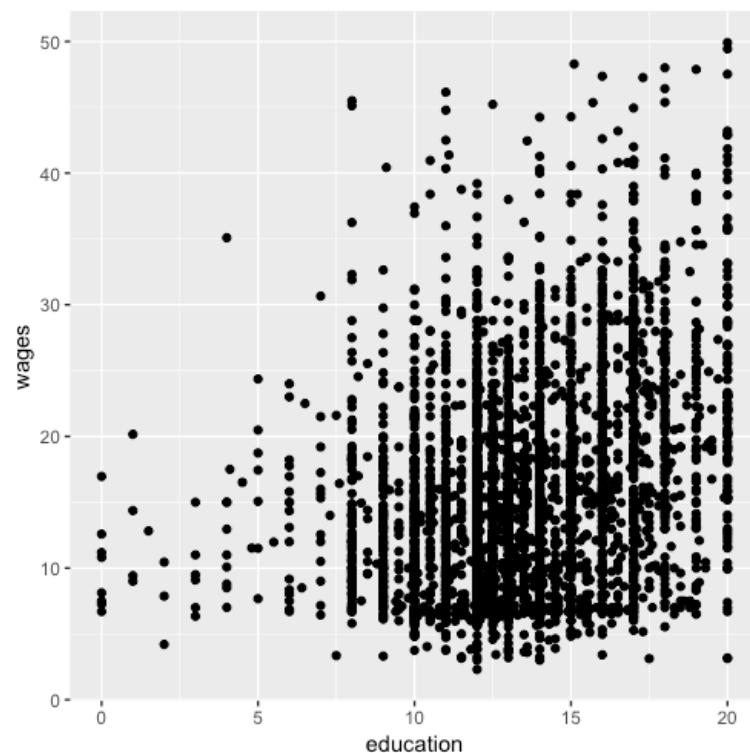
Statistics

Facets

Geometries

Aesthetics

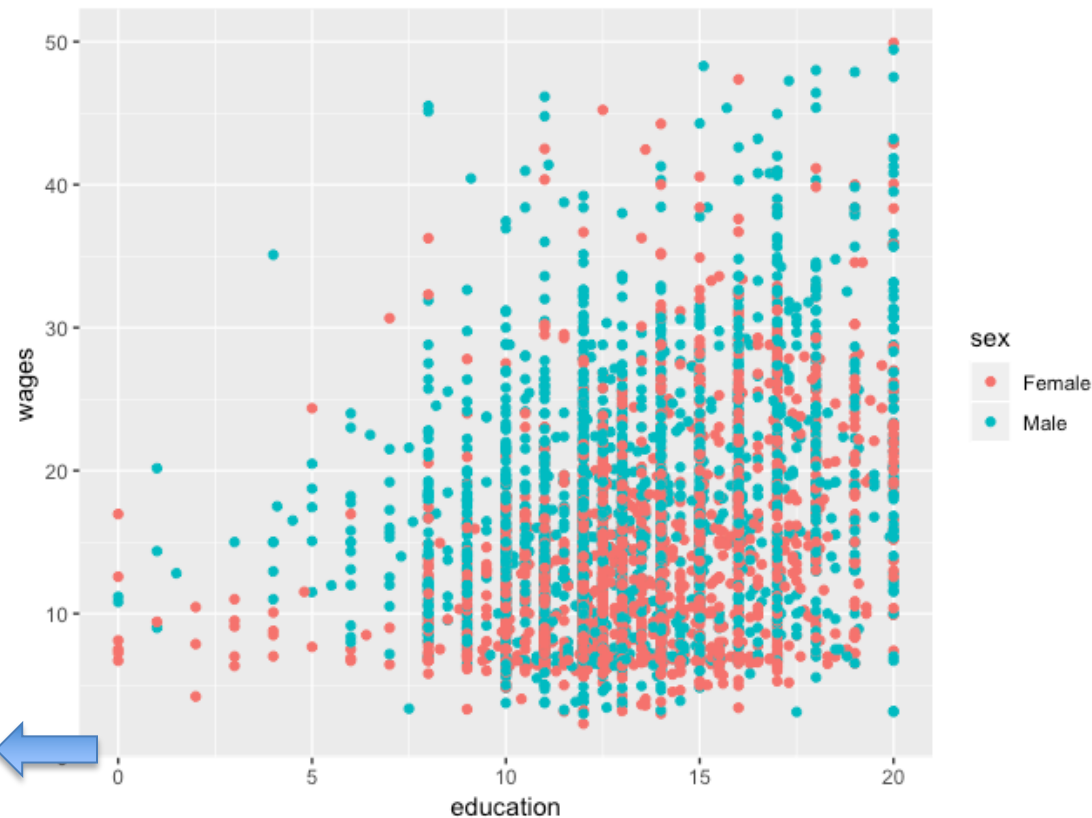
Data





# aesthetic (again)

```
g <- ggplot(data=SLID)
g <- g + geom_point(
  aes(x=education, y=wages, colour=sex))
g
```



Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

Statistics

Facets

Geometries

Aesthetics

Data

# geometry (again)

```
g <- ggplot(data=SLID)
g <- g + geom_smooth(
  aes(x=education, y=wages, colour=sex))
g
```

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

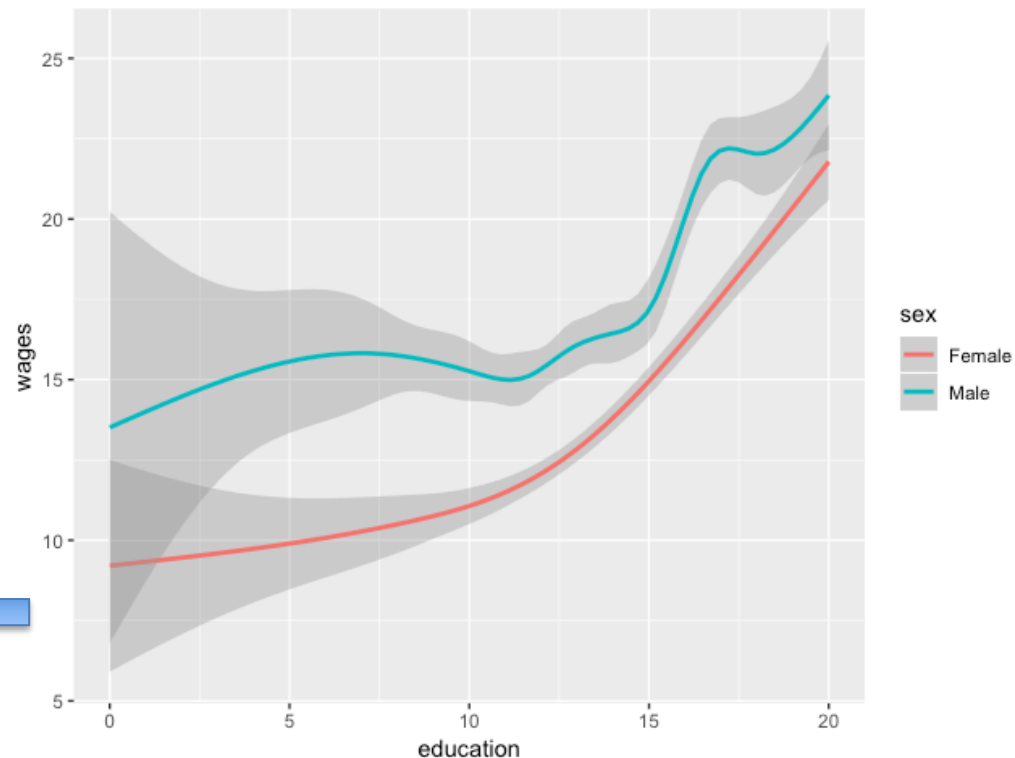
Statistics

Facets

Geometries

Aesthetics

Data



# aesthetic

- Exercises: combine points and smooth lines

# facet

```
g <- ggplot(data=SLID,  
            aes(x=education, y=wages)) +  
  geom_smooth()  
g <- g + facet_wrap(~sex)  
g
```

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

Statistics

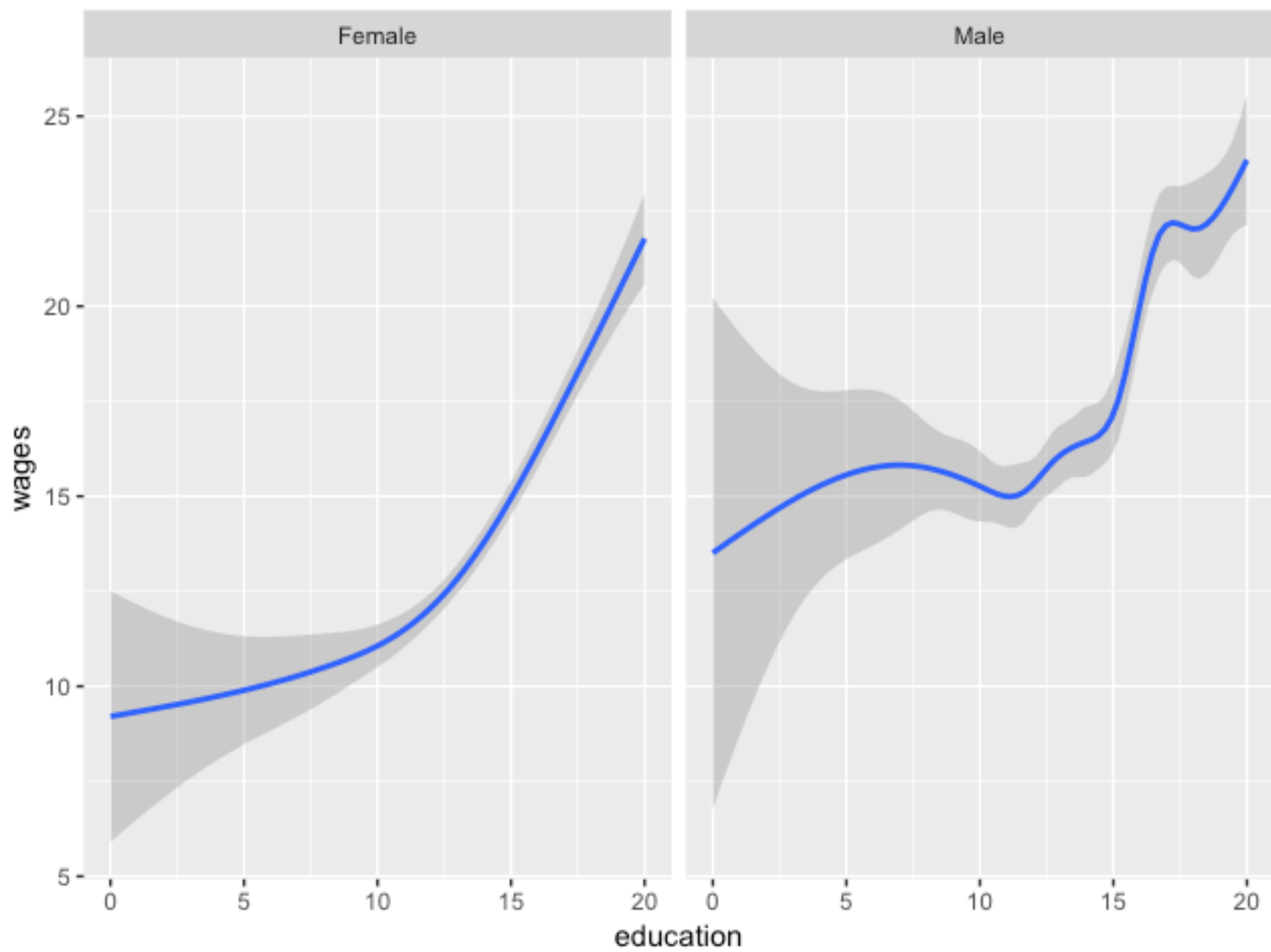
Facets

Geometries

Aesthetics

Data





# facet

- Exercise:
  - try other variables
  - try `facet_grid(var1 ~ var2)`

# statistics

```
g = ggplot(  
  data.frame(country=c("UK", "USA"), population=c(66.02, 325.7)) ) +  
  geom_bar(aes(x=country, y=population), stat="identity")  
g
```

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

Statistics

Facets

Geometries

Aesthetics

Data



# statistics

```
g = ggplot(data=SLID)
g = g + geom_bar(aes(x=sex, y=wages),
  stat="summary", fun.y = "mean")
g
```

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

Statistics

Facets

Geometries

Aesthetics

Data





# statistics

- Exercises:
  - try other statistics (mean, max, count)
  - try adding group/fill/colour to aesthetic

# coordinates

- `coord_map()`
- Exercises:
  - try other `coord_flip()`, `coord_equal()`

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme

Coordinates

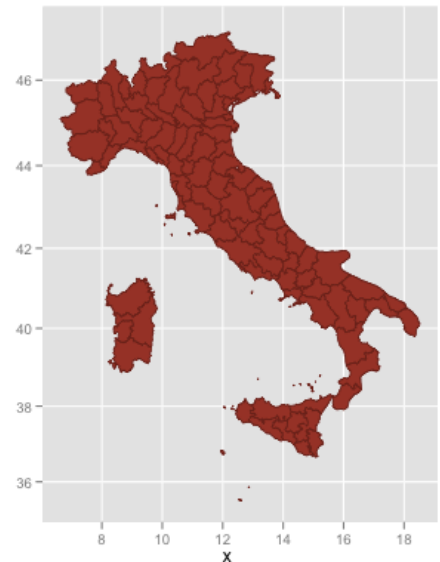
Statistics

Facets

Geometries

Aesthetics

Data



# themes

- Exercises:
  - Find a list of themes online, try a few different themes.

```
g + theme(  
  text = element_text(family = "Open Sans", size = 11),  
  plot.margin = unit(c(44.0,11.0,43.0,11.0), "points"),  
  plot.background = element_rect(fill = "#ecf7f1", size = 0),  
  panel.background = element_blank(),  
  panel.grid.major.x = element_line(color = "#c0c0c0", size = 0.2),  
  panel.grid.major.y = element_blank(),  
  panel.grid.minor = element_blank(),  
  axis.ticks = element_blank(),  
  axis.title.x = element_blank(),  
  axis.title.y = element_blank(),  
  legend.position="none"
```

Describes all the non-data ink

Plotting space for the data

Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent the data

Scales onto which data is mapped

The actual variables to be plotted

Theme



Coordinates

Statistics

Facets

Geometries

Aesthetics

Data

# export

```
ggsave("filename.pdf", g, width=12, height=4)
```

# Where to look for help?

- <https://ggplot2.tidyverse.org/reference/>
- <http://sape.inf.usi.ch/quick-reference/ggplot2/>
- Built-in help (?function)
- StackOverflow / Google: always include “ggplot2” as a search term