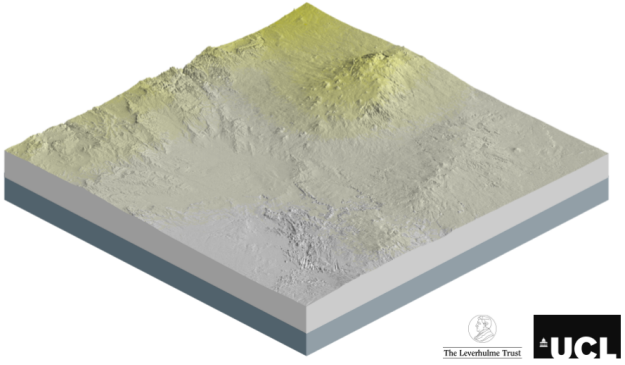
R for Archaeologists

Visualisation

Modelling prehistoric hunting strategies in the eastern Levant, 23–8 kya



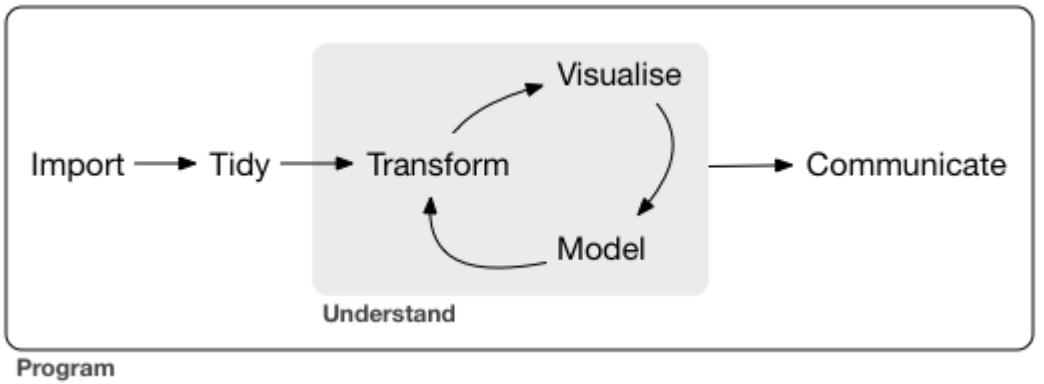
Joe Roe
Research assistant
University of Copenhagen

https://joeroe.io
https://github.com/joeroe
https://twitter.com/joeroe90

Day	Time	Topics	Exercises
Wednesday	**10–13**	**Introduction to visualisation with R** • Visualisation in data science • Overview of R graphics packages • Plotting with `base`	*Quantitative Archaeology*, ch. 5
	14–17	**Tidy visualisation with `ggplot2`** • The 'grammar of graphics' • Visualisation and tidy data • Plotting with `ggplot2`	*R for Data Science*, ch. 3 & 7
Thursday	**10–13**	**Advanced visualisations for archaeologists** • Radiocarbon chronologies with 'ridgeplots' • Cartography in R: `sf` and `ggplot2` • 3D topography with `rayshader`	To be provided
	14–17	**Publication-ready visualisations** • What makes an effective graphic? • Advanced `ggplot2`: aesthetics and themes • Publishing with `RMarkdown`	*R for Data Science*, ch. 27–28

Visualisation with R

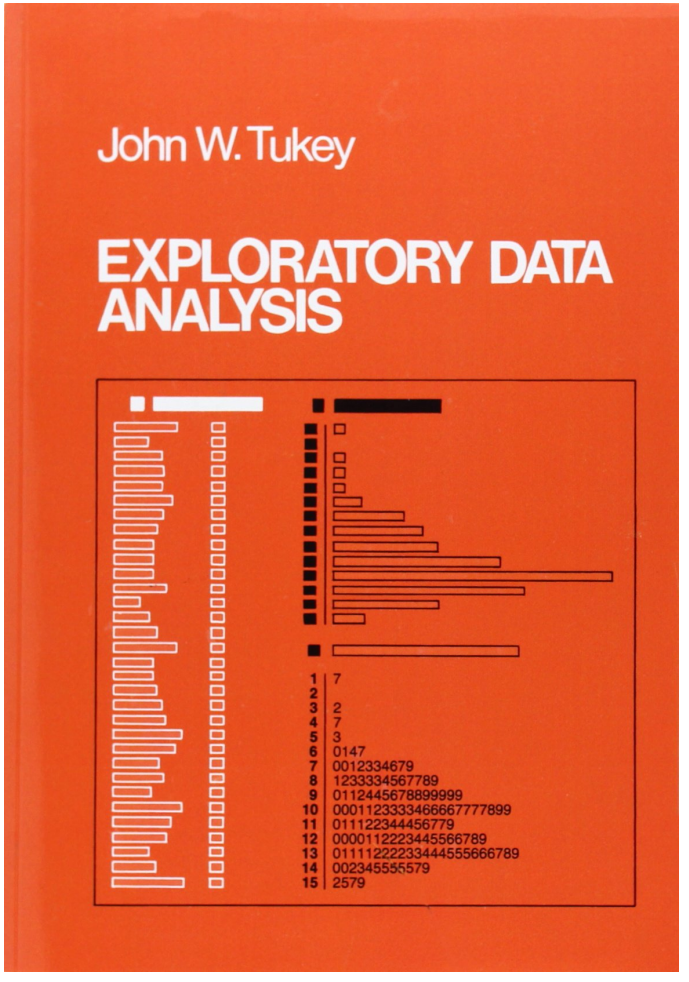
Introduction



Data science workflow. After Wickham & Grolemund, *R for Data Science*

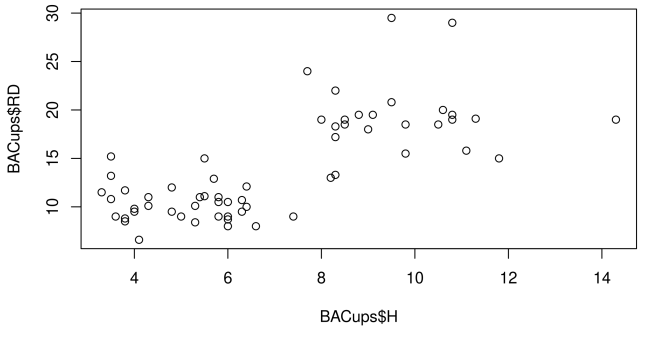
Exploratory Data Analysis (EDA)

> Exploratory data analysis (EDA) is an essential step in any research analysis. The primary aim with exploratory analysis is to examine the data for distribution, outliers and anomalies to direct specific testing of your hypothesis. It also provides tools for hypothesis generation by visualizing and understanding the data usually through graphical representation. EDA aims to assist the natural patterns recognition of the analyst.
>
> *Komorowski et al. 2016*

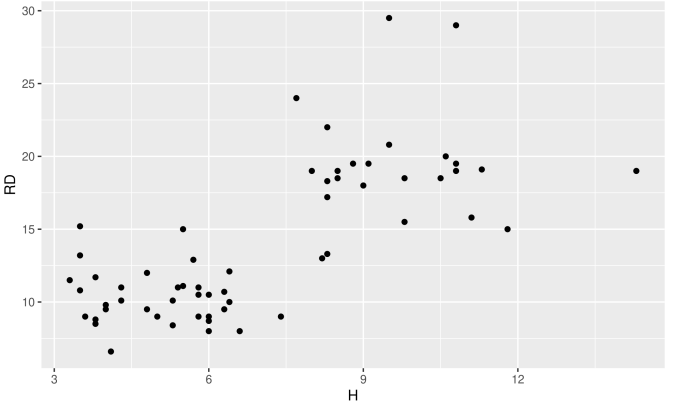


Graphics packages in R

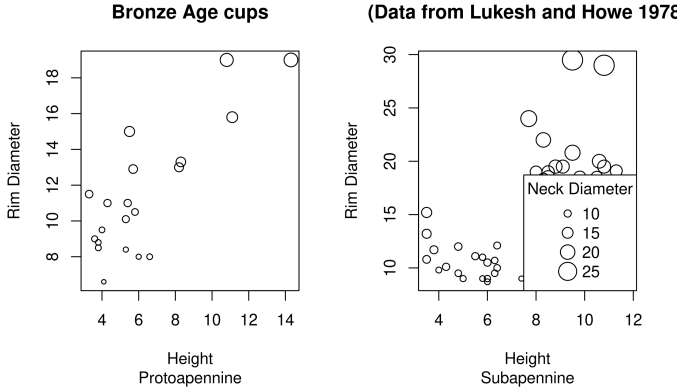
```
library("archdata")
data("BACups")
```

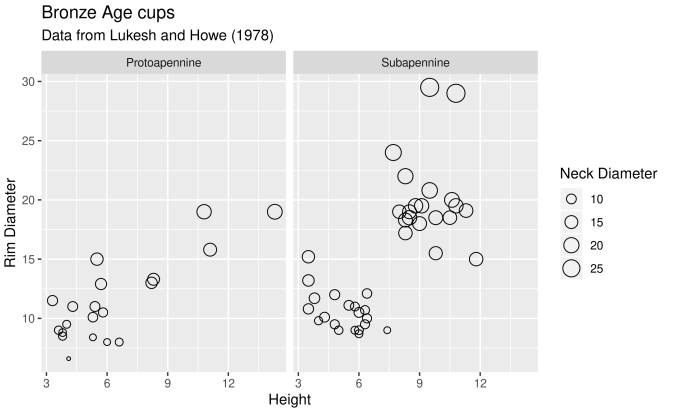



```
plot(BACups$H, BACups$RD)
```




```
ggplot(BACups, aes(x = H, y = RD)) +
  geom_point()
```






```
par(mfrow = c(1,2))
plot(BACups[BACups$Phase=="Protoapenr
     BACups[BACups$Phase=="Protoapenr
     cex = BACups[BACups$Phase=="Prot
     main = "Bronze Age cups",
     sub = "Protoapennine",
     xlab = "Height",
     ylab = "Rim Diameter")
plot(BACups[BACups$Phase=="Subapennir
     BACups[BACups$Phase=="Subapennir
     cex = BACups[BACups$Phase=="Suba
     main = "(Data from Lukesh and Ho
     sub = "Subapennine",
     xlab = "Height",
     ylab = "Rim Diameter")
legend("bottomright",
       title = "Neck Diameter",
       pch = 1,
       legend = c(10, 15, 20, 25),
       pt.cex = c(1, 1.5, 2, 2.5))
```

```
ggplot(BACups, aes(x = H, y = RD, siz
  geom_point(shape = 1) +
  facet_wrap(vars(Phase)) +
  labs(x = "Height",
       y = "Rim Diameter",
       size = "Neck Diameter",
       title = "Bronze Age cups",
       subtitle = "Data from Lukesh a
```

Plotting with base: exercises

Carlson 2017, *Quantitative Methods in Archaeology Using R*

Chapter 5 – Looking at Data: Graphs (skip the first section)

- Create a new script in RStudio
- Replicate all Carlson's code
- Think about Carlson's code – how could you adapt it to do something different?
- Answer questions as comments (`# Here is my answer...`)
- Shout at me if you get stuck!

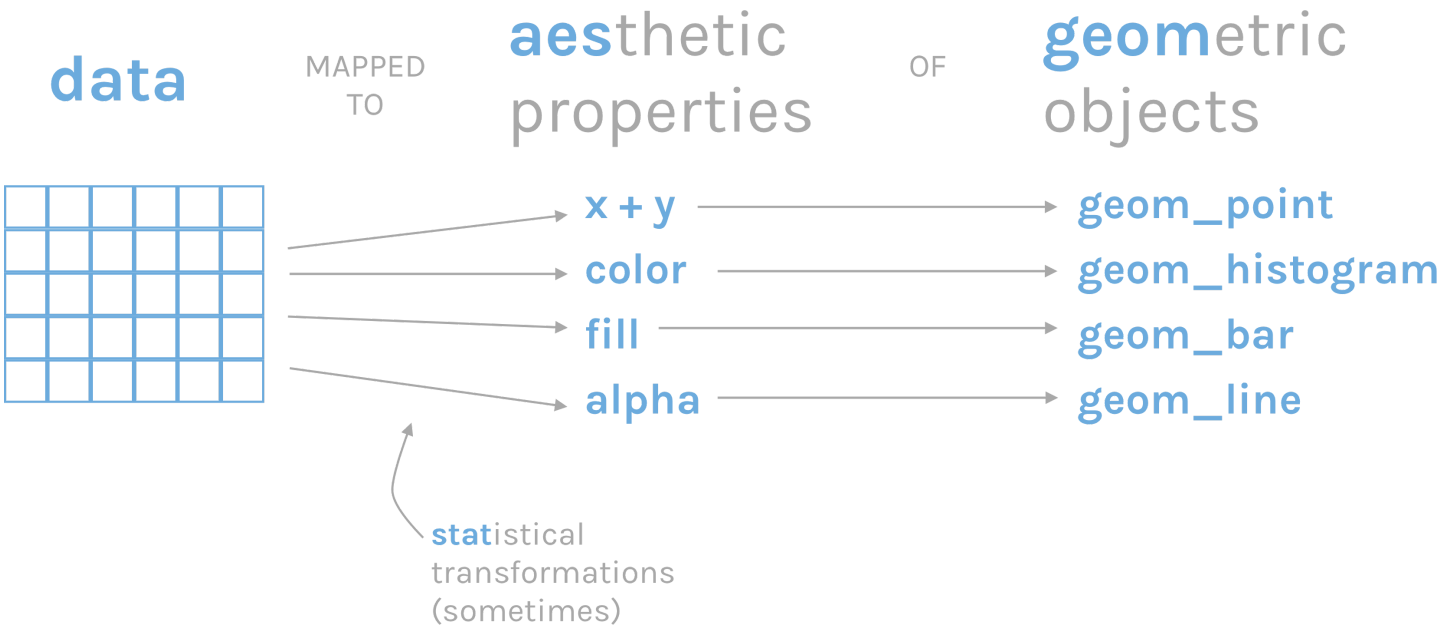
Plotting with base: extension exercises

1. Using the `MaskSite` data, make a) a pie chart and b) a bar plot of the frequency of artefacts by category. Which do you think is more effective?

Tidy Visualisation

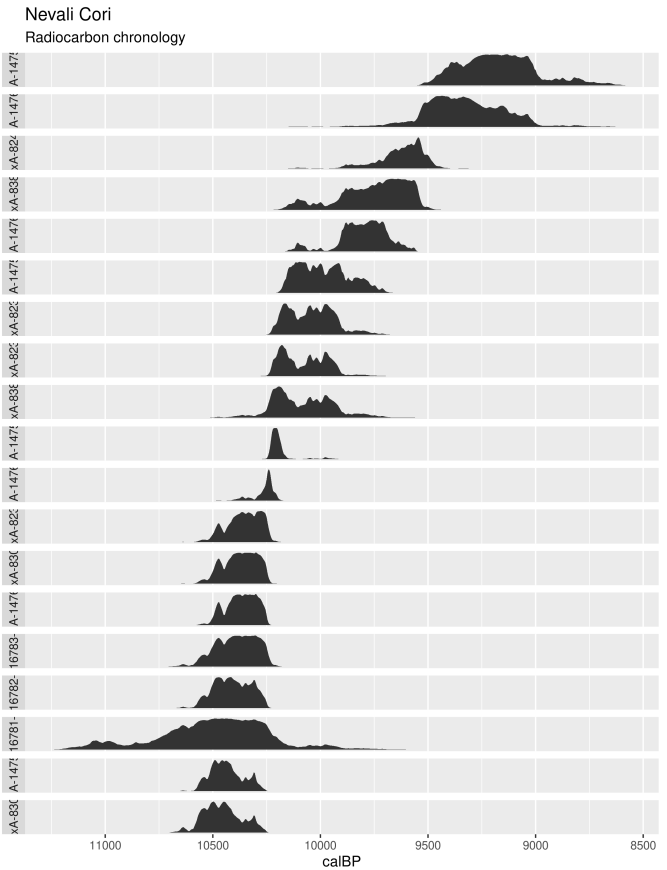
The `ggplot2` package

The Grammar of Graphics



ggplot2 and tidy data

```
data("emedyd") # From rcarbon
emedyd %>%
  filter(SiteName == "Nevali Cori") %
  arrange(CRA) %>%
  mutate(LabID = as_factor(LabID)) %>
  mutate(CalDate = cal(CRA, Error, ve
  unnest_legacy() %>%
  ggplot(aes(calBP, PrDens)) +
  facet_wrap(~LabID, ncol = 1,
             scales = "free_y",
             strip.position = "left")
  geom_area() +
  scale_x_reverse() +
  labs(title = "Nevali Cori",
      subtitle = "Radiocarbon chronol
  theme(axis.title.y = element_blank(
        axis.text.y = element_blank()
        axis.ticks.y = element_blank(
        panel.grid.minor.y = element_
        panel.grid.major.y = element_
```



Plotting with `ggplot2`: exercises

Wickham & Grolemund, *R for data science*

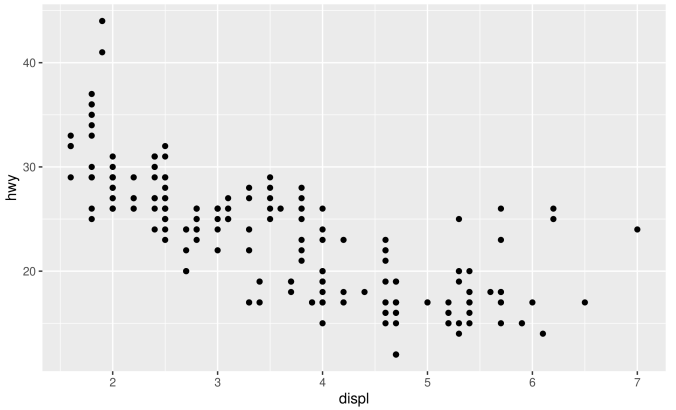
https://r4ds.had.co.nz/

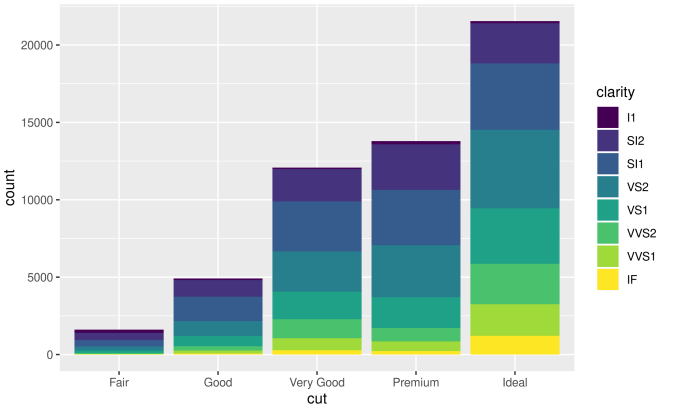
Chapter 2: *Data visualisation*

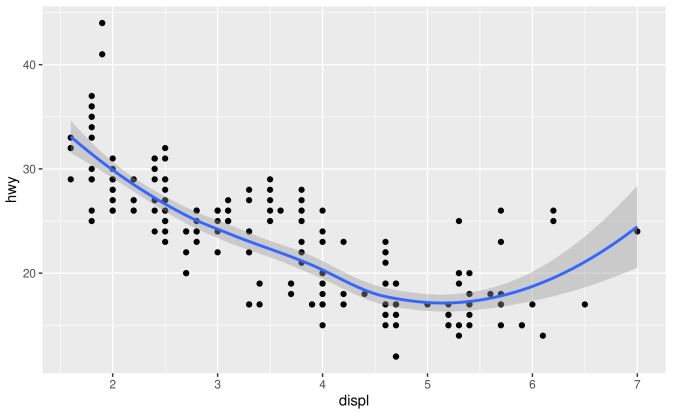
Chapter 7: *Exploratory Data Analysis* (if you have time)

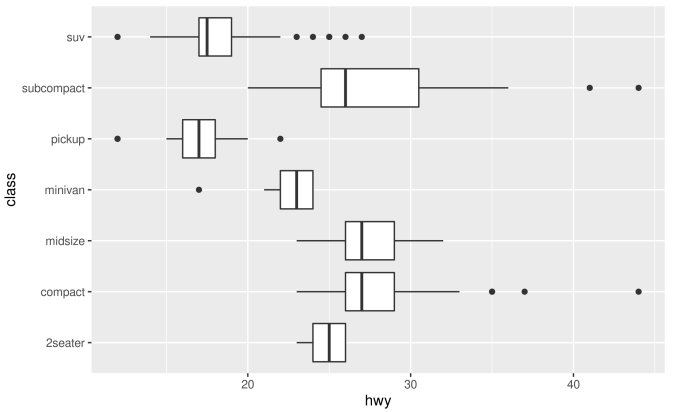
- Install the `tidyverse` metapackage
- Run Hadley's code examples
- Answer questions as comments (`# Here is my answer...`)
- Do the 'exercises' sections – they're very informative!

Plotting with `ggplot2`: applications?







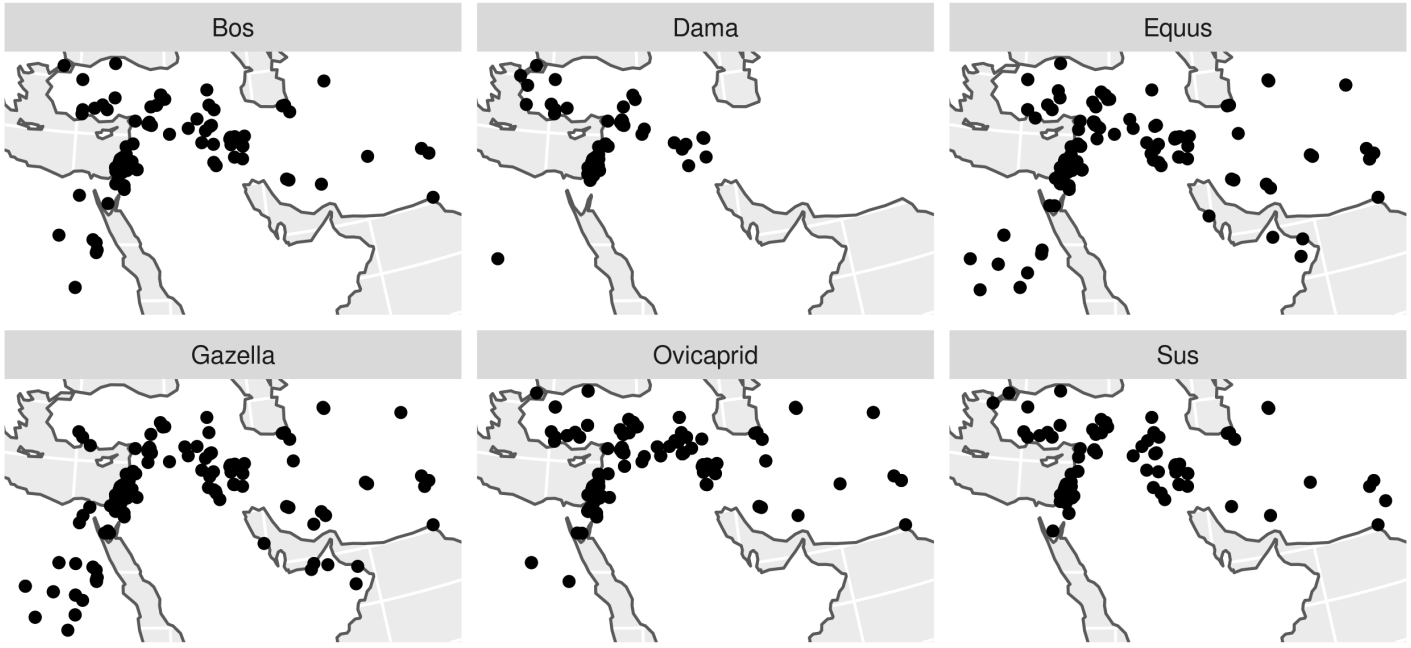


Advanced visualisations

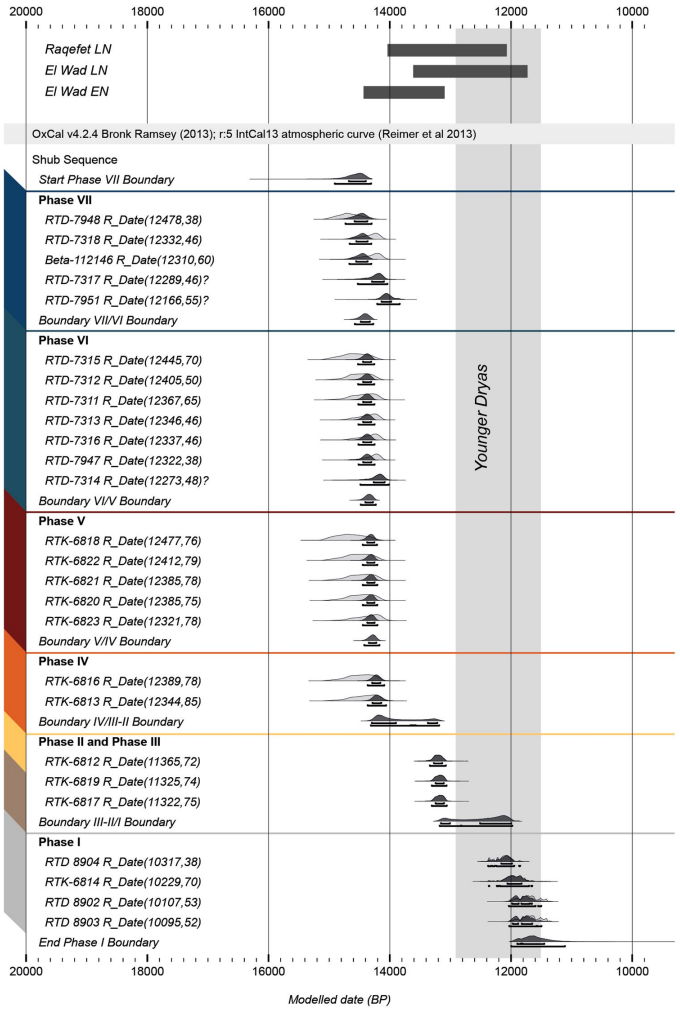
Plotting archaeological data

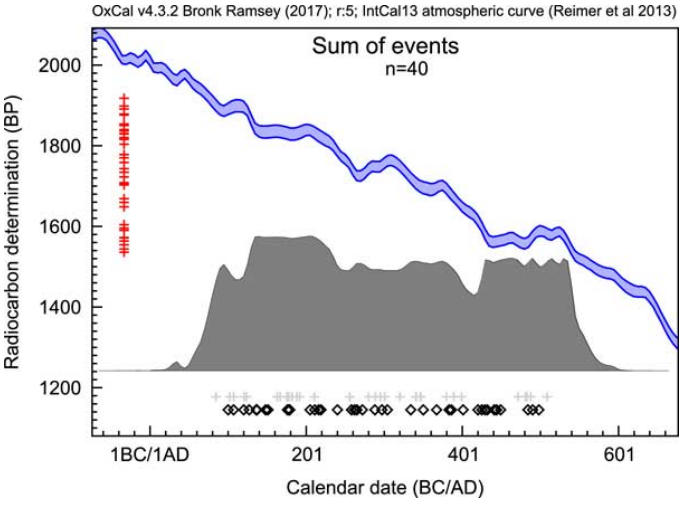
Cartography: R as a GIS

- Spatial data
  - `rgdal`, `rgeos`
  - `sp`, `sf`
  - `raster`, `stars`
- Geostatistics
  - `gstat`, `spdep`, `nlme`
  - `landsat`, `dismo`
  - Many, many more...
- Base maps + open data
  - `maps`, `maptools`, `rnaturalearth`
  - `ggmap`, `leaflet`, `mapview`
  - `rOpenSci`
- Cartography
  - `ggplot2`
  - `rasterVis`


Bos
Dama
Equus
Gazella
Ovicaprid
Sus

Radiocarbon





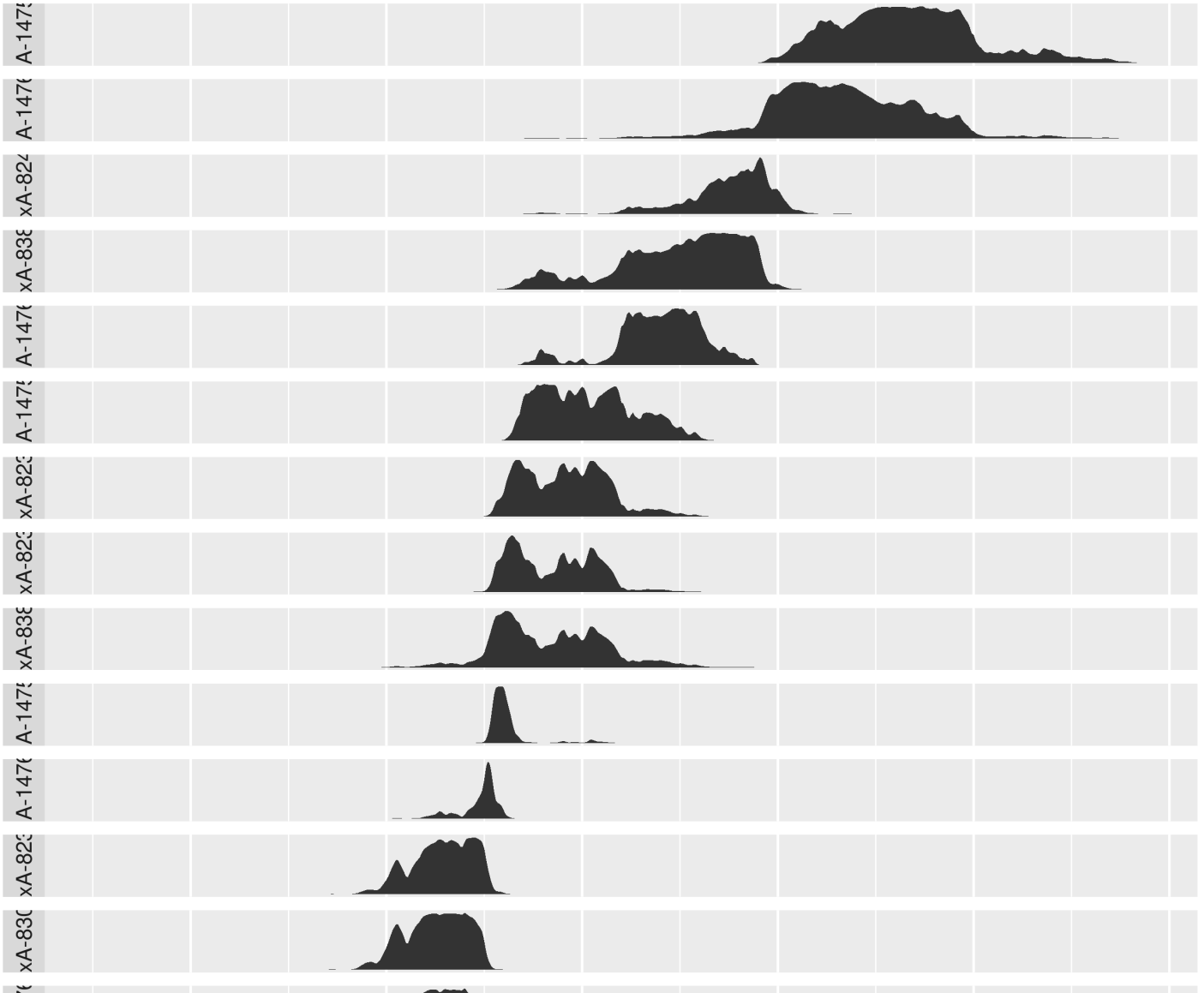
Bronk Ramsey 2017

Radiocarbon packages in R

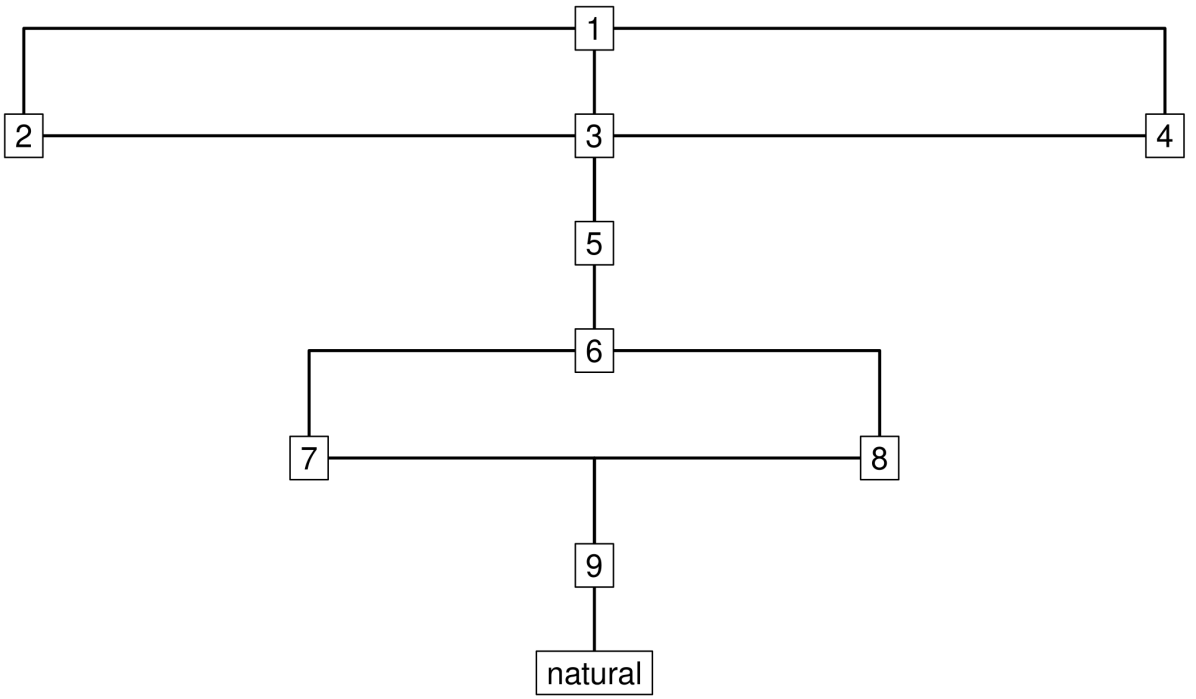
- Databases
  - `c14bazAAR`
- Calibration
  - `oxcAAR` (interface for OxCal)
  - `rcarbon` (including summed distributions)
  - `stratigraphr` (WIP!)

Nevali Cori

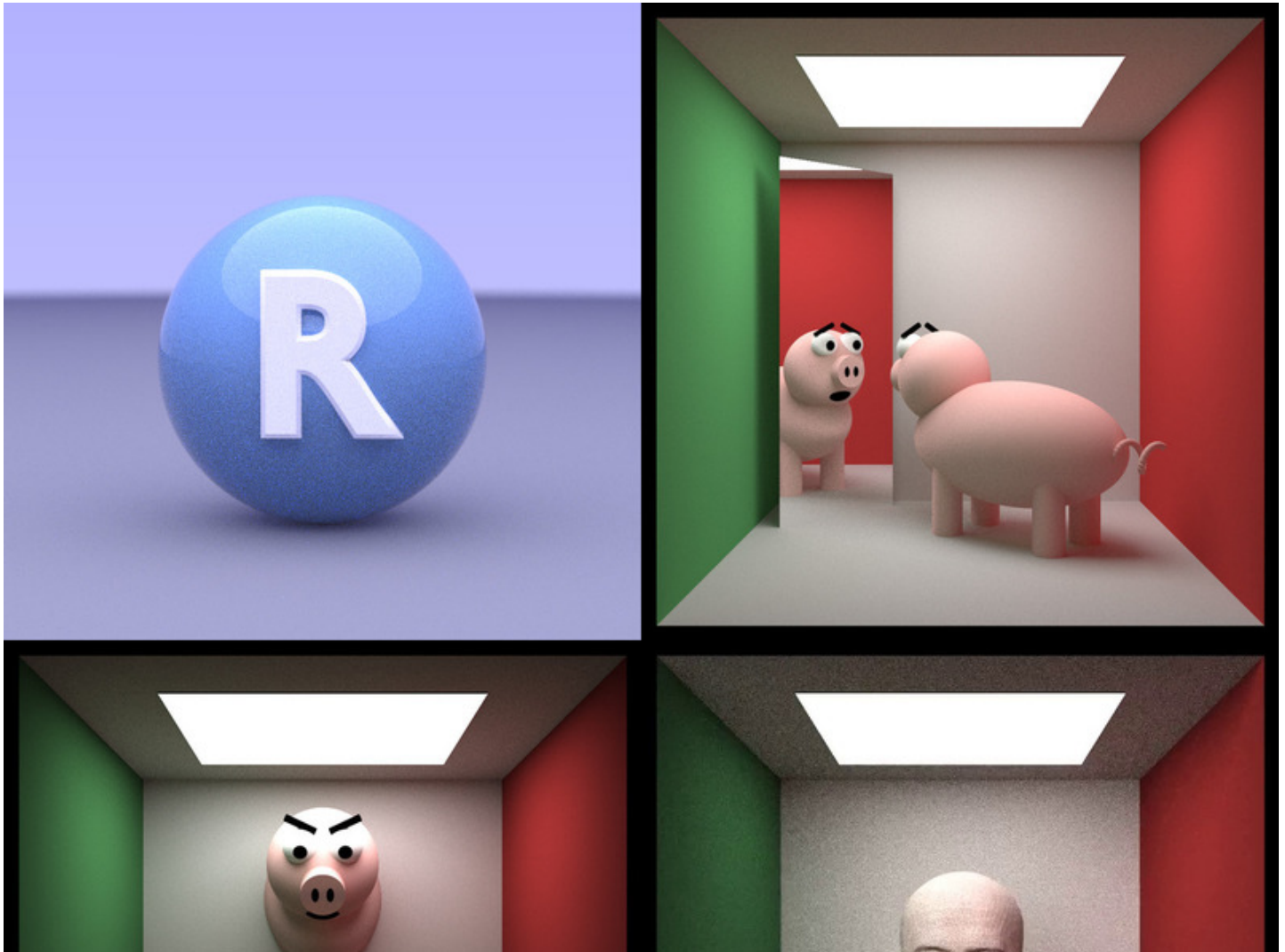
Radiocarbon chronology

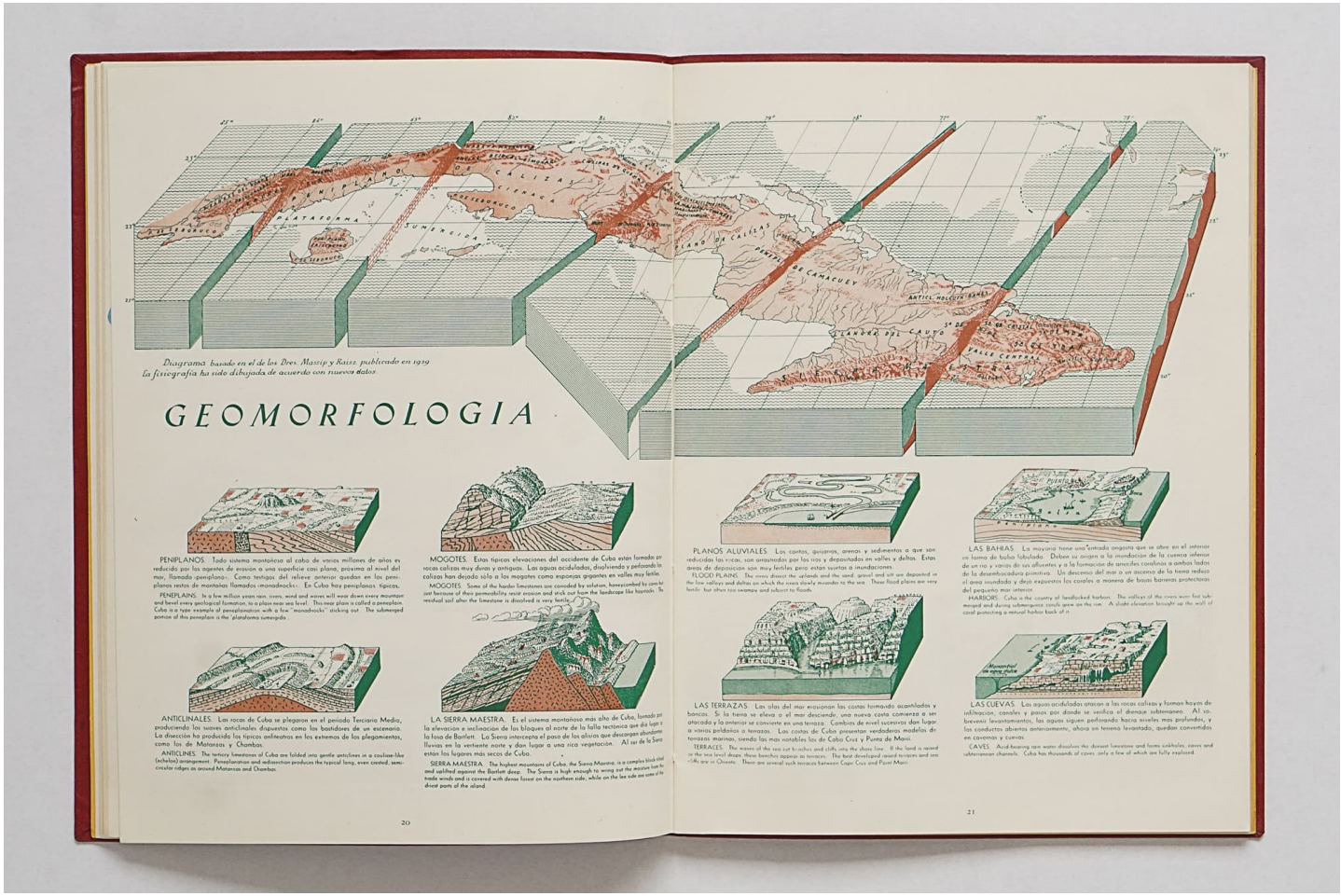


Stratigraphy (tidygraph & ggraph)

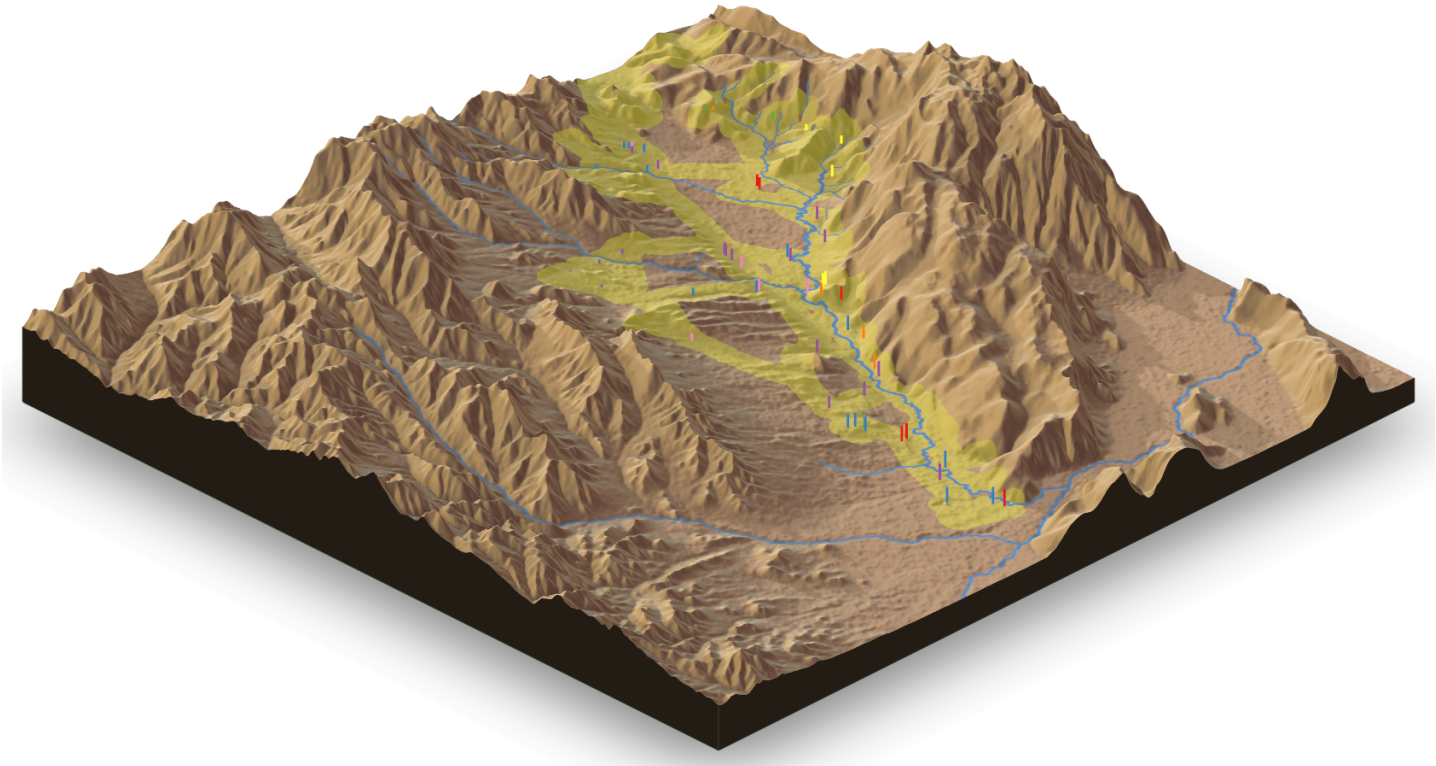


3D (rayshader & rayrender)





Gerardo Canet & Erwin Raisz, *Atlas de Cuba* (1949)



Publication-ready visualisations

ggplot2 and RMarkdown

Publishing with RMarkdown

- Notebooks
- Papers & theses (`rrtools`)
- Books (`bookdown`, e.g. *R for Data Science*)
- Slides (`xaringan`, e.g. these!)
- Interactive interfaces (`shiny`, e.g. https://joeroe.shinyapps.io/LAC2018_fieldwalkr/)

RMarkdown for science: research compendia

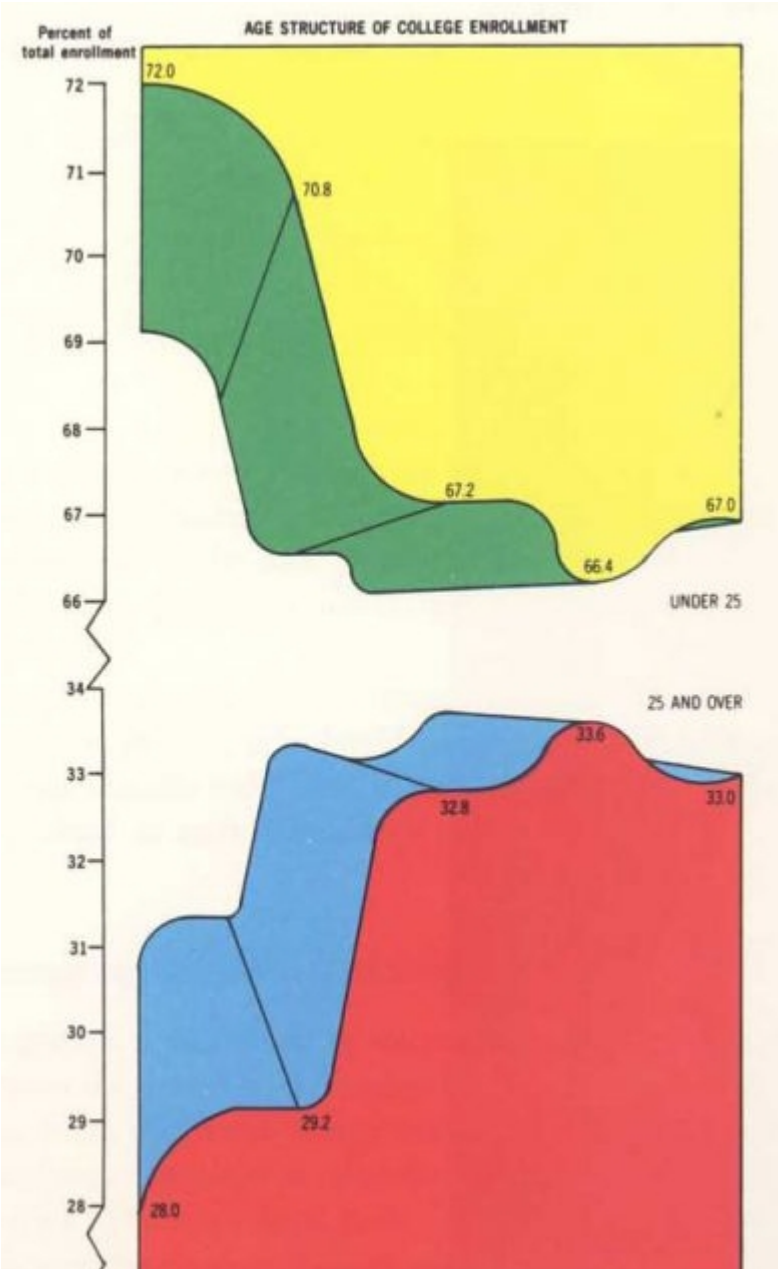
A **research compendium** packages together all the **data**, **code**, **figures**, and **text** that comprise a scientific analysis in a single place (e.g. an R package).

The idea is to make the analysis fully **reproducible** from start to finish.

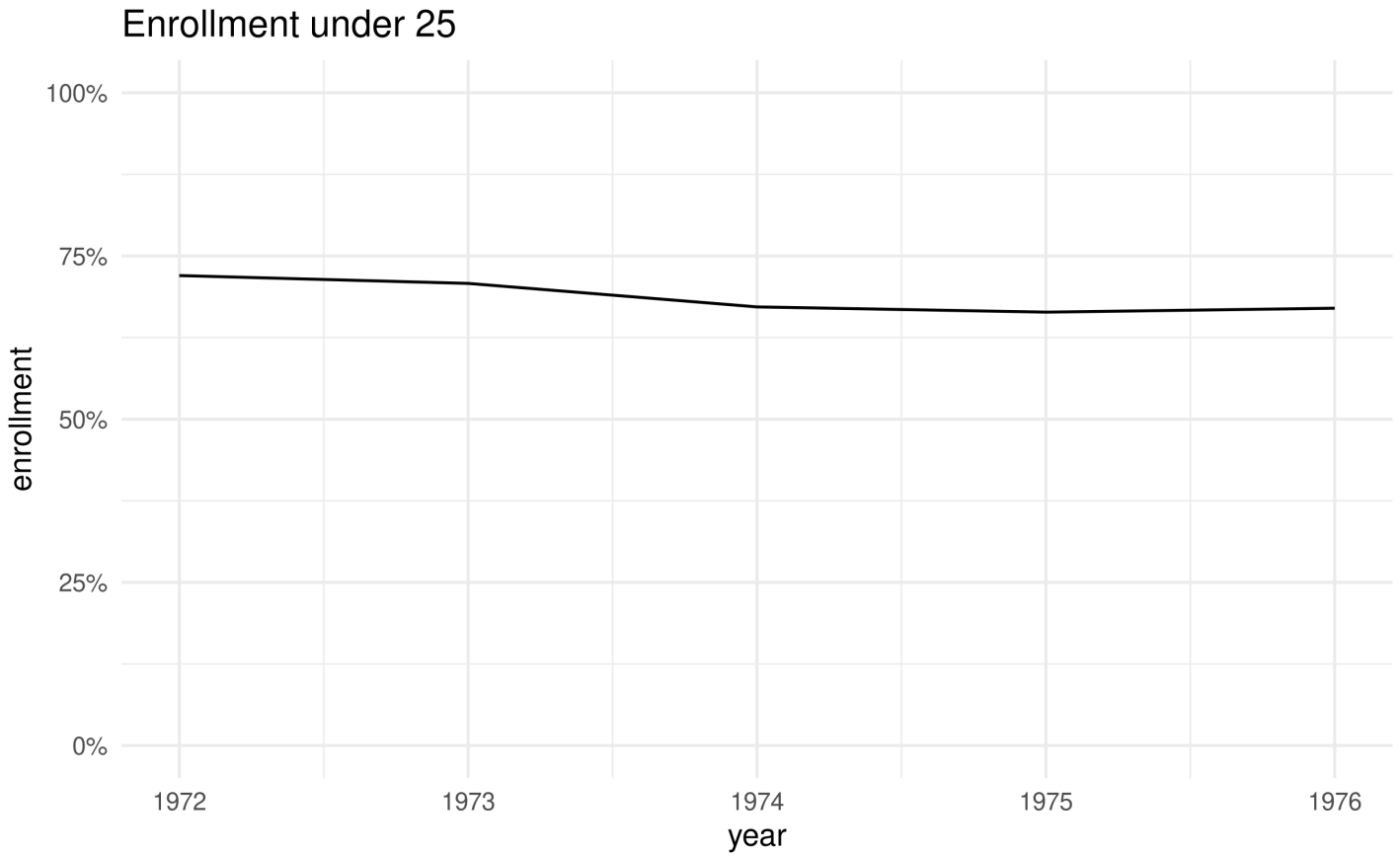
The R package `rrtools` assists with creating these in R.

Effective visualisations: Tufte's principles

1. Above all else show data.
2. Maximize the data-ink ratio.
3. Erase non-data-ink.
4. Erase redundant data-ink.
5. Revise and edit.


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

```
tufte_data %>%
  filter(age_group == "<25") %>%
  ggplot(aes(x = year, y = enrollment)) +
  geom_line() +
  scale_y_continuous(limits = c(0, 1), labels = percent) +
  labs(title = "Enrollment under 25") +
  theme_minimal()
```



`ggplot2` and `RMarkdown`: exercises

Wickham & Grolemund, *R for data science*

https://r4ds.had.co.nz/

Chapter 27: *R Markdown*

Chapter 28: *Graphics for communication*

- Run Hadley's code examples
- Answer questions as comments (`# Here is my answer...`)
- Do the 'exercises' sections – they're very informative!

Further reading

- Wickham & Grolemund, *R for Data Science*, https://r4ds.had.co.nz/
  - Hadley also recommends lots of good books!
- Wickham 2009, *ggplot2: Elegant Graphics for Data Analysis*
- Xie 2015, *Dynamic Documents with R and knitr*
- Marwick et al. 2017, *Packaging Data Analytical Work Reproducibly Using R (and Friends)*, https://doi.org/10.1080/00031305.2017.1375986
- Wickham 2020, *Mastering `Shiny`*, https://mastering-shiny.org/
- RStudio cheatsheets, https://rstudio.com/resources/cheatsheets/