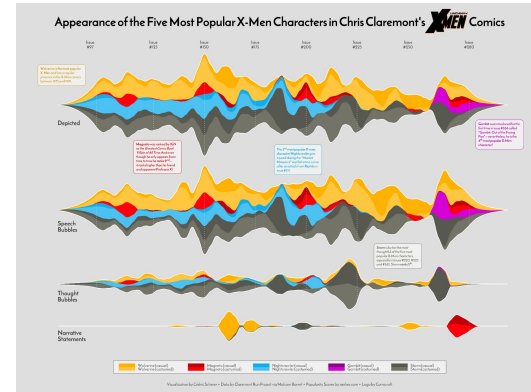# Elaboration of real data

PhDToolbox R coding final workshop

Aims and Goals

1. Choose one dataset
2. Explore it
3. Wrangle your data (if needed)
4. Make gorgeous visualizations of data
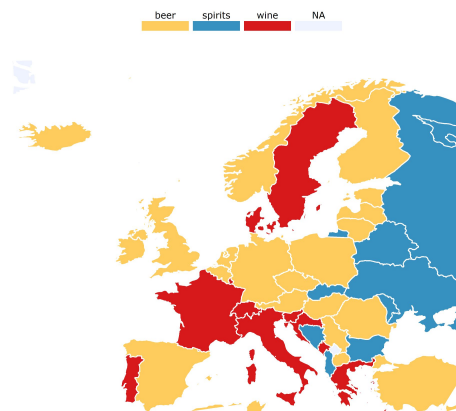5. Explain to the other teams your results with a 20' presentation

- The use of functions not explained by us is recommended
- Use multiple plug-in of the ggplot package is a warm suggestion
- Amaze us!

Appearance of the Five Most Popular X-Men Characters in Chris Claremont's XMEN Comics

Average Percent Bee Colony Lost Per Year
2018   2019

Preferred alcoholic beverages - Europe
For countries with at least 10 total servings per capita per year

beer   spirits   wine   NA

Dataset 1

You can find all details at the following link:

https://www.davidzeleny.net/anadat-r/doku.php/en:data:aravo

A dataset of a real vegetation survey on the French Alps that include vegetation data (Braun-Blanchet method), species traits and environmental variables.

You can indulge in cross-referencing source data and find traces of ecological patterns between species, community composition and environmental variables.

Combining the fourth-corner and the RLQ methods for assessing trait responses to environmental variation

Stéphane Dray ✉, Philippe Choler, Sylvain Dolédec, Pedro R. Peres-Neto, Wilfried Thuiller, Sandrine Pavoine, Cajo J. F. ter Braak

Environmental variables

Code	Variable
Aspect	Relative south aspect (opposite of the sine of aspect with flat coded 0)
Slope	Slope inclination (°)
Form	Microtopographic landform index [1 (convexity); 2 (convex slope); 3 (right slope); 4 (concave slope); 5 (concavity)]
Snow	Mean snowmelt date (Julian day) averaged over 1997–1999
PhysD	Physical disturbance, i.e., percentage of unvegetated soil due to physical processes
ZoogD	Zoogenic disturbance, i.e., quantity of unvegetated soil due to marmot activity [no; some; high]

Species attributes (traits)

Code	Trait
Height	Vegetative height (cm)
Spread	Maximum lateral spread of clonal plants (cm)
Angle	Leaf elevation angle estimated at the middle of the lamina
Area	Area of a single leaf
Thick	Maximum thickness of a leaf cross section (avoiding the midrib)
SLA	Specific leaf area
Nmass	Mass-based leaf nitrogen content
Seed	Seed mass

Dataset 2

Haunted Places in the United States

Ready to create the most ghosty data visualization ever?

variable	class	description
city	character	The city where the place is located.
country	character	The country where the place is located (always "United States")
description	character	A text description of the place. The amount of detail in these descriptions is highly variable.
location	character	A title for the haunted place.
state	character	The US state where the place is located.
state_abbrev	character	The two-letter abbreviation for the state.
longitude	double	Longitude of the place.
latitude	double	Latitude of the place.
city_longitude	double	Longitude of the city center.
city_latitude	double	Latitude of the city center.

plot your maps!

One dataset, hundreds of possible visualization
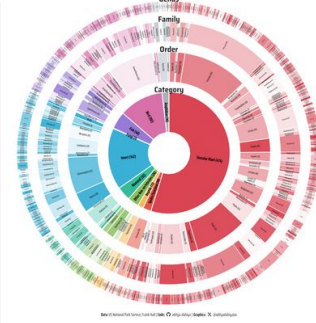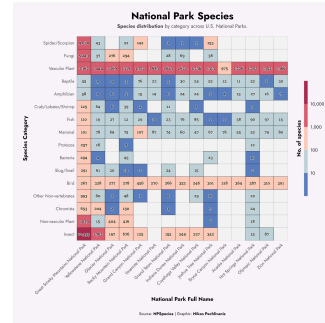
Dataset 3

National Park Species

A species inventory of the 15 most visited National Parks in USA

1) Clean your data as suggested by authors

2) Tidy-up your data and enjoy data visualization stuff!

Dataset 4

Are you Ready for It?

There are three main datasets.

1. The first is taylor_album_songs, which includes lyrics and audio features from the Spotify API for all songs on Taylor's official studio albums. Notably this excludes singles released separately from an album (e.g., Only the Young, Christmas Tree Farm, etc.), and non-Taylor-owned albums that have a Taylor-owned alternative (e.g., Fearless is excluded in favor of Fearless (Taylor's Version)). We stan artists owning their own songs.
2. You can access Taylor's entire discography with taylor_all_songs. This includes all of the songs in taylor_album_songs plus EPs, individual singles, and the original versions of albums that have been re-released as Taylor's Version.
3. Finally, there is a small data set, taylor_albums, summarizing Taylor's album release history.

You are free to explore the most futile aspects of modern pop, but never say it to Marco Gamba.