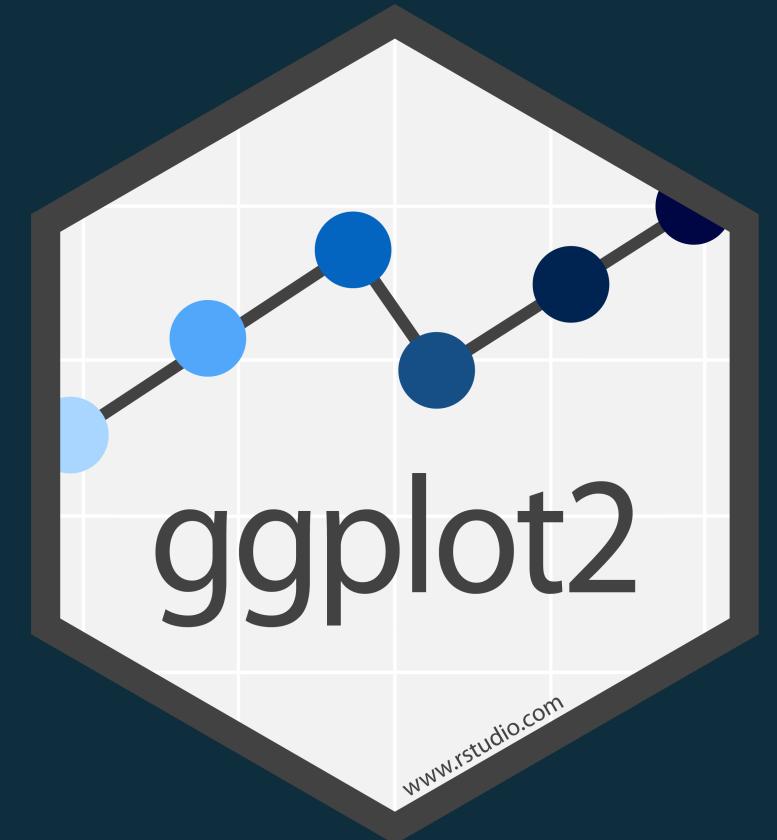


Data visualization with ggplot2

Introduction to R - Day 2

Instructor: Selina Baldauf

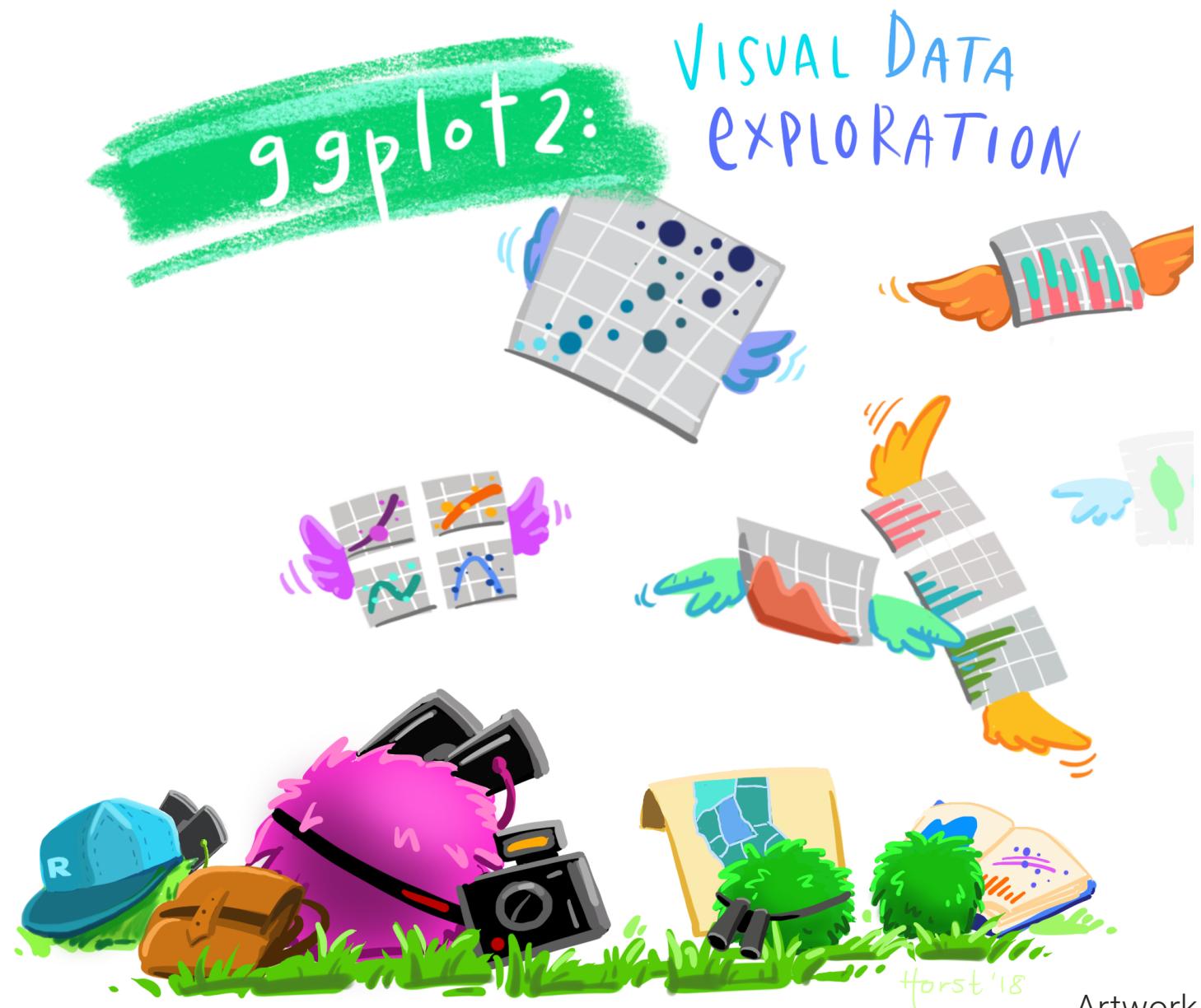
Freie Universität Berlin - Theoretical Ecology



2021-06-15 (updated: 2021-08-02)

ggplot2

- ggplot2 was developed by Hadley Wickham in 2005
- most popular package for building graphics in R
- one of the most downloaded packages from CRAN



ggplot2:

Build a data
MASTERpiece



HORST '18

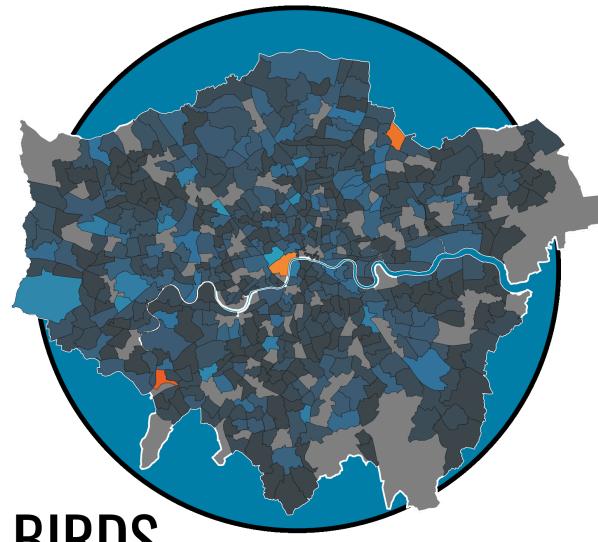
Artwork by Allison Horst 4 / 15

A ggplot showcase

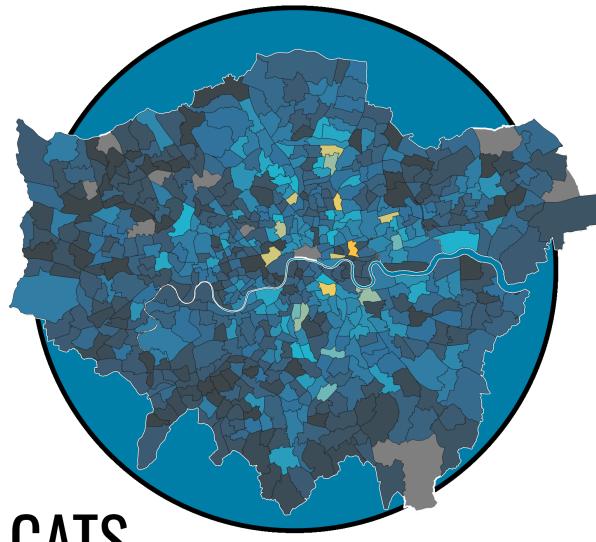
Some examples of plots you can create with ggplot

Frequency of Rescues of Birds, Cats and Dogs in London from 2009-2021

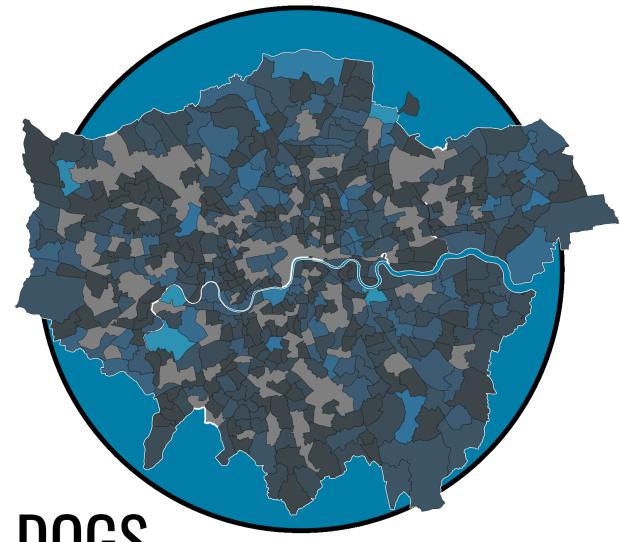
Illustrated below in three choropleth maps are rescues of birds, cats and dogs in London wards. Darker colors indicate lower rescue numbers while brighter colors indicate a greater number of rescues in that ward.



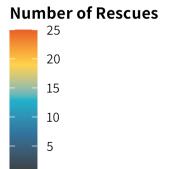
BIRDS



CATS



DOGS

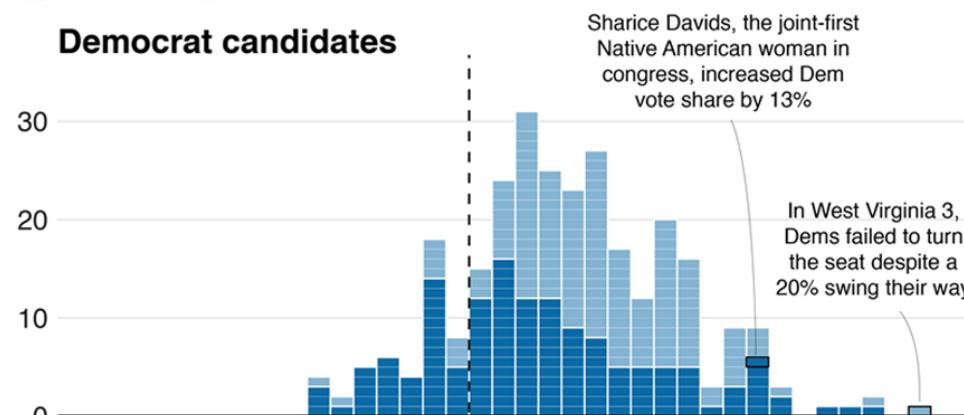


Data: London.gov | Graphic: @jakekaupp

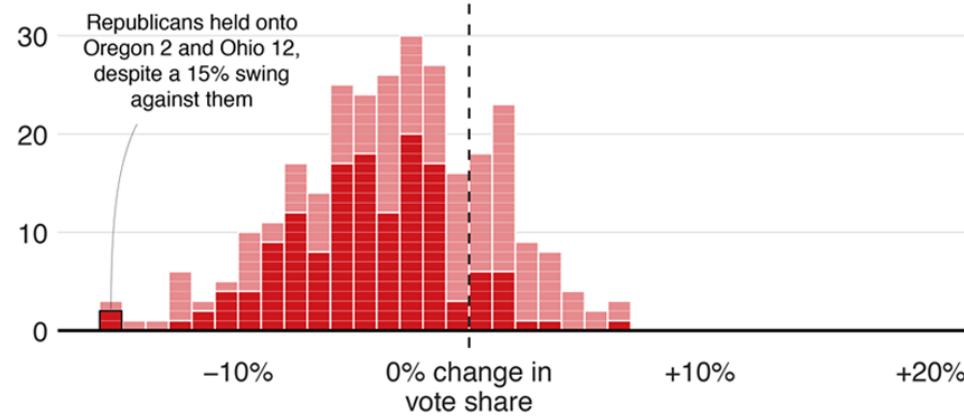
Blue wave

■ Won seat ■ Didn't win

Democrat candidates

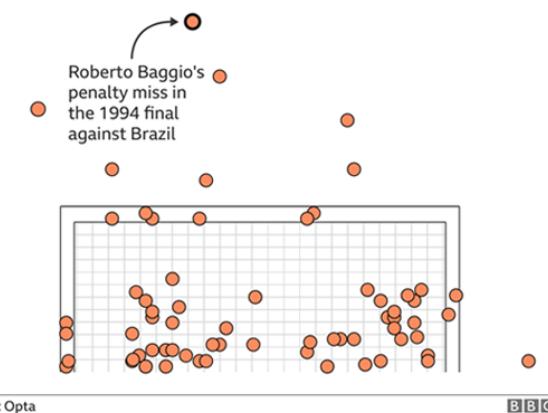


Republican candidates



Where penalties are saved

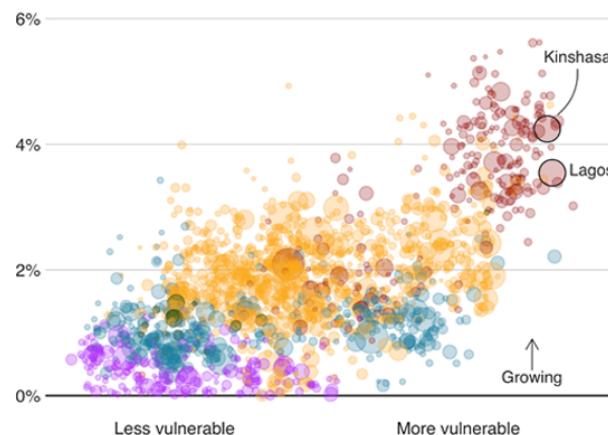
World Cup shootout misses and saves, 1982-2014



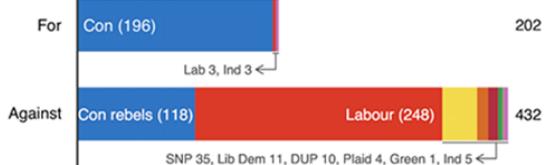
Fast-growing cities face worse climate risks

Population growth 2018-2035 over climate change vulnerability

■ Africa ■ Asia ■ Americas ■ Europe ■ Oceania

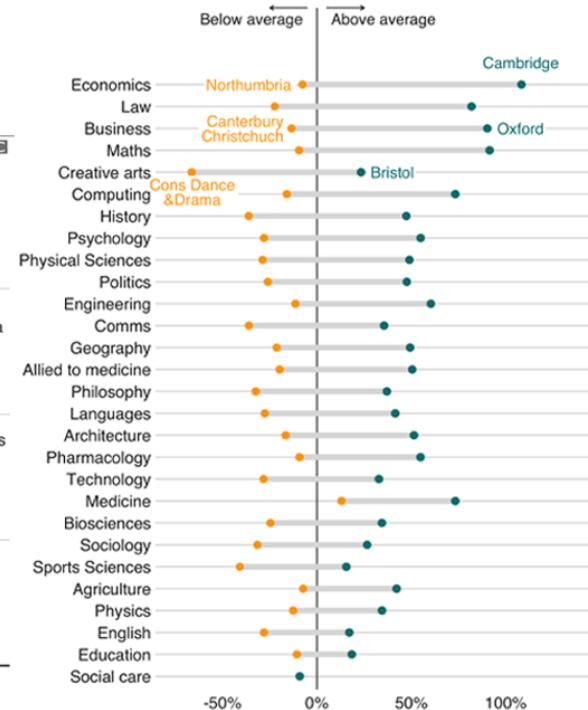


MPs rejected Theresa May's deal by 230 votes

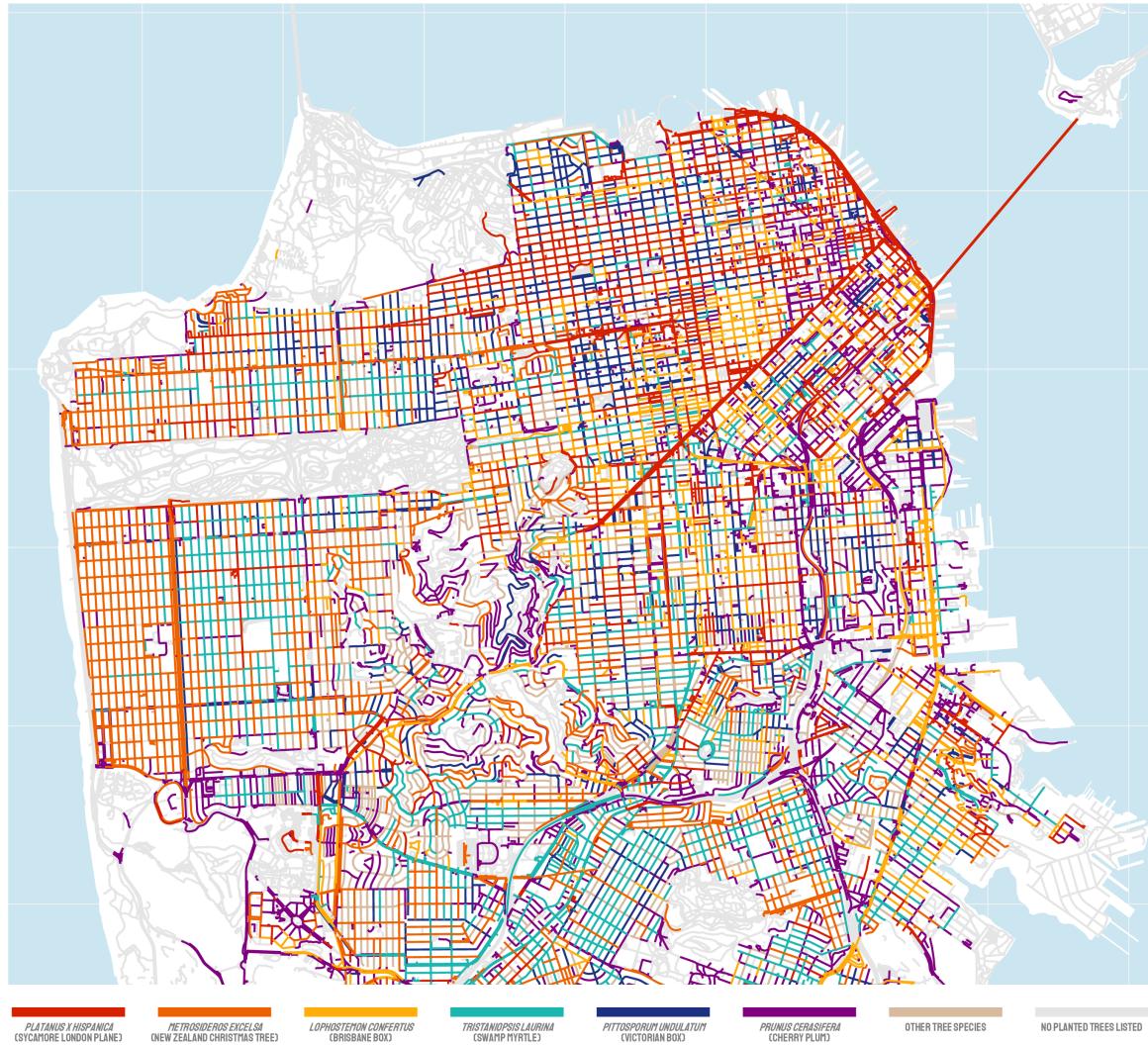


Earnings vary across unis even within subjects

Impact on men's earnings relative to the average degree

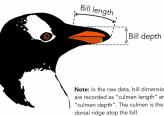


THE DOMINANT TREE SPECIES' PLANTED ALONG SAN FRANCISCO'S ROADS

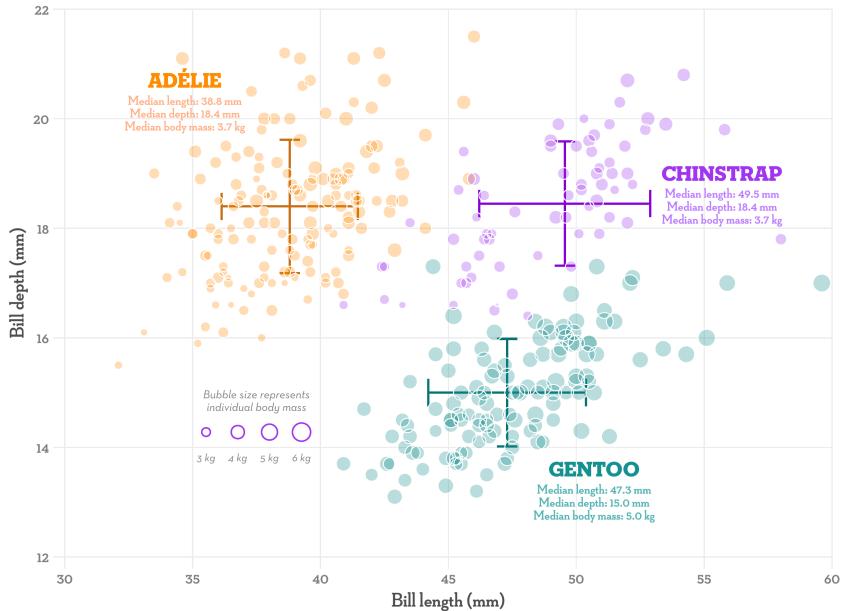


BILL DIMENSIONS OF BRUSH-TAILED PENGUINS

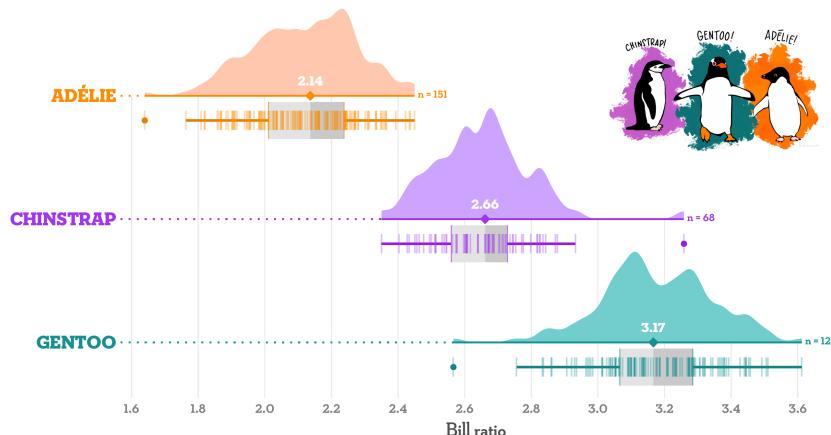
Pygoscelis adeliae (Adélie penguin) • *P. antarctica* (Chinstrap penguin) • *P. papua* (Gentoo penguin)



A. Scatterplot of bill length versus bill depth (error bars show median +/- sd)



B. Distribution of the bill ratio, estimated as bill length divided by bill depth



Note: In the original data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal (upper) ridge of a bird's bill.
Visualization: Cédric Scherer • Data: Gorman, Williams & Fraser (2014) DOI: 10.1371/journal.pone.0099081 • Illustrations: Allison Horst

Visualization by Cédric Scherer,
code available on [Github](#)

Advantages of ggplot

- consistent grammar/structure
- flexible structure allows you to produce any type of plot
- highly customizable appearance (themes)
- many extension packages that provide
 - Additional plot types
 - Additional themes
 - Color palettes
 - Animation
 - Composition of multiple plots
 - ...
- Active community that provides help and inspiration

Basic idea of ggplot

Shorten the distance from mind to page

(Hadley Wickham)

`ggplot2` is an implementation of the grammar of graphics by Leland Wilkinson

- coherent system for describing and building graphs
- basic idea: stack distinct **layers of graphical elements** to create a plot

Layers of a ggplot

Layer	Function	Details
Data	<code>ggplot(data)</code>	The data that you want to plot.
Mapping	<code>aes()</code>	Aesthetic mappings of the geometric and statistical objects.
Geometries	<code>geom_*</code>	Graphical representation of aesthetics as point, line, polygon, ...

Layers of a ggplot

Layer	Function	Details
Data	<code>ggplot(data)</code>	The data that you want to plot.
Mapping	<code>aes()</code>	Aesthetic mappings of the geometric and statistical objects.
Geometries	<code>geom_*</code> ()	Graphical representation of aesthetics as point, line, polygon, ...
Scales	<code>scale_*</code> ()	Translate between variable values and properties (e.g. categories -> colour)
Coordinates	<code>coord_*</code> ()	Interprets and maps the position values of the data.
Facets	<code>facet_*</code> ()	Split your plot into multiples.
Themes	<code>theme()</code> and <code>theme_*</code> ()	Visual look of the plot not related to the data.

Summary I

The `ggplot` package has a built-in data set called `msleep` about the sleep times of 83 mammals.

Data variables (among others):

- `vore: carnivore, omnivore, herbivore?
- `conservation: conservation status of the animal
- `sleep_total: total amount of sleep [h]
- `brainwt: brain weight [kg]
- `bodywt: body weight [kg]

If you want to know more about the data set:

```
library(ggplot2)
str(msleep)
?msleep
```

The `ggplot()` function initializes a ggplot object. Every ggplot needs this function.

Now you

Task 1: Create your own penguin ggplots

Find the task description [here](#)