Introduction to Infographics and Data Visualization (Spring 2023)

20231 - JMM622-O

Final Project

Yuri Souza

Project Two-JMM622-O

Yuri Souza

2023-04-30

This file contains the workflow I used for the Atlantic forest project

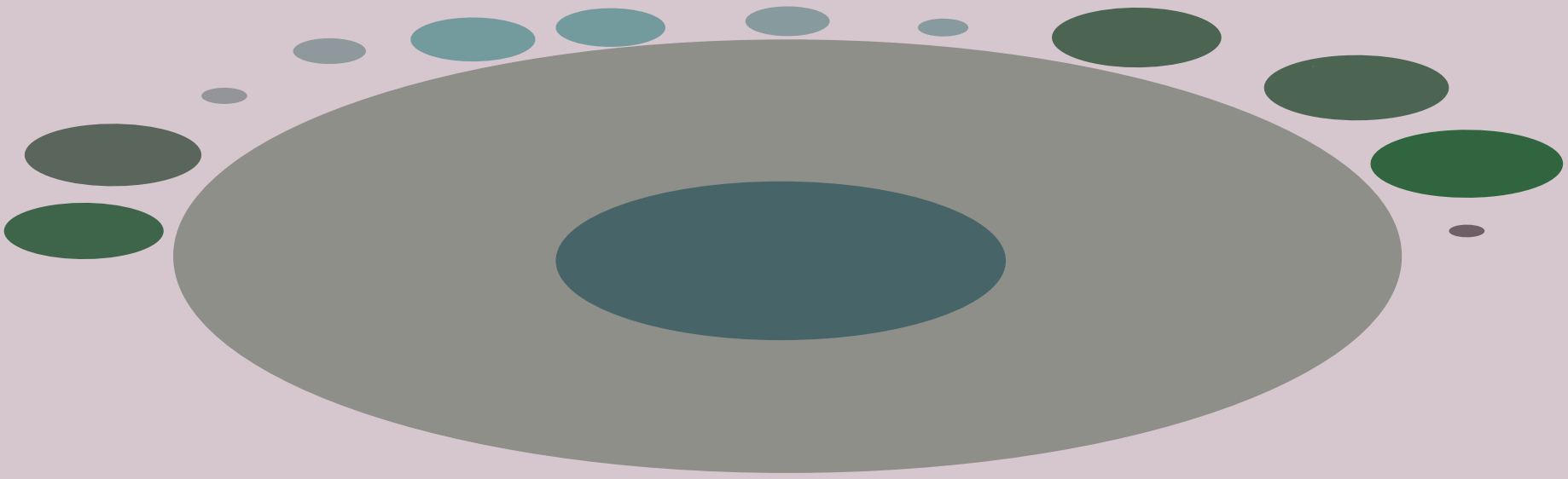
First Version

I choose to work with Atlantic Forest and the relationship of numbers of peer reviewed papers and the number of species. I picked up this topic because over time i have seen that some species have received more attention than others. Everything I used here is also available on my *GitHub*.

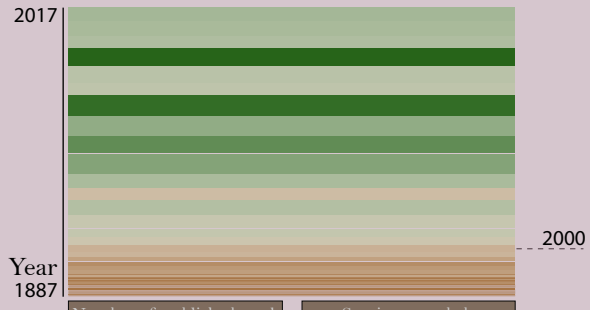
Version 1. Since the beginning, I wanted to create this infographic using more stylish, trying to show my data as it was a tree. The reason was to get more attention to the topic beyond the use of charts and graphics.I first started by creating a sketch of the size and the number of boxes that I would like to include using a landscape infographic.

But then I realized the long orientation would work better. Then, I got the data using the Datapapers published in the scientific *Ecology* journal. These Datapapers contain all the available data about specific species. For the purpose of this project, I decided to use only the published data due to their robustness and integrity. I tried to create a tree using the *canopy* to represent the loss of the forest cover by the Brazilian state using a **bubble chart**. Under this chart, I put a **heat map chart** relating the number of published papers per year since the first published paper. The reason I chose this chart was totally based on design. I wanted something to represent the *trunk* of the tree. I knew that there would have a pattern from more papers being published recently than in the past, so I used the brown color as the one to represent years with few publications. Thus, I could use this color to start the *root* part of the tree. For the *root* part, I used the **Sankey diagram chart** with the purpose of splitting the published papers into the number of publications per group of animals and the number of recorded species by these papers per year. My idea was to show that the organisms diversity is the root of the *tropical forest*. By using the **Sankey diagram**, I wanted to show that some species have been studied more than others and that there was a relationship between the number of studies and the number of species recorded by year. Basically, the more work we do, the more species we record. For small animal groups, it seems that we haven't reached the S-Curve or the expected number of species. I used **linear charts** to represent the number of published papers and the recorded species over time, linking it to their respective *Sankey curve*.

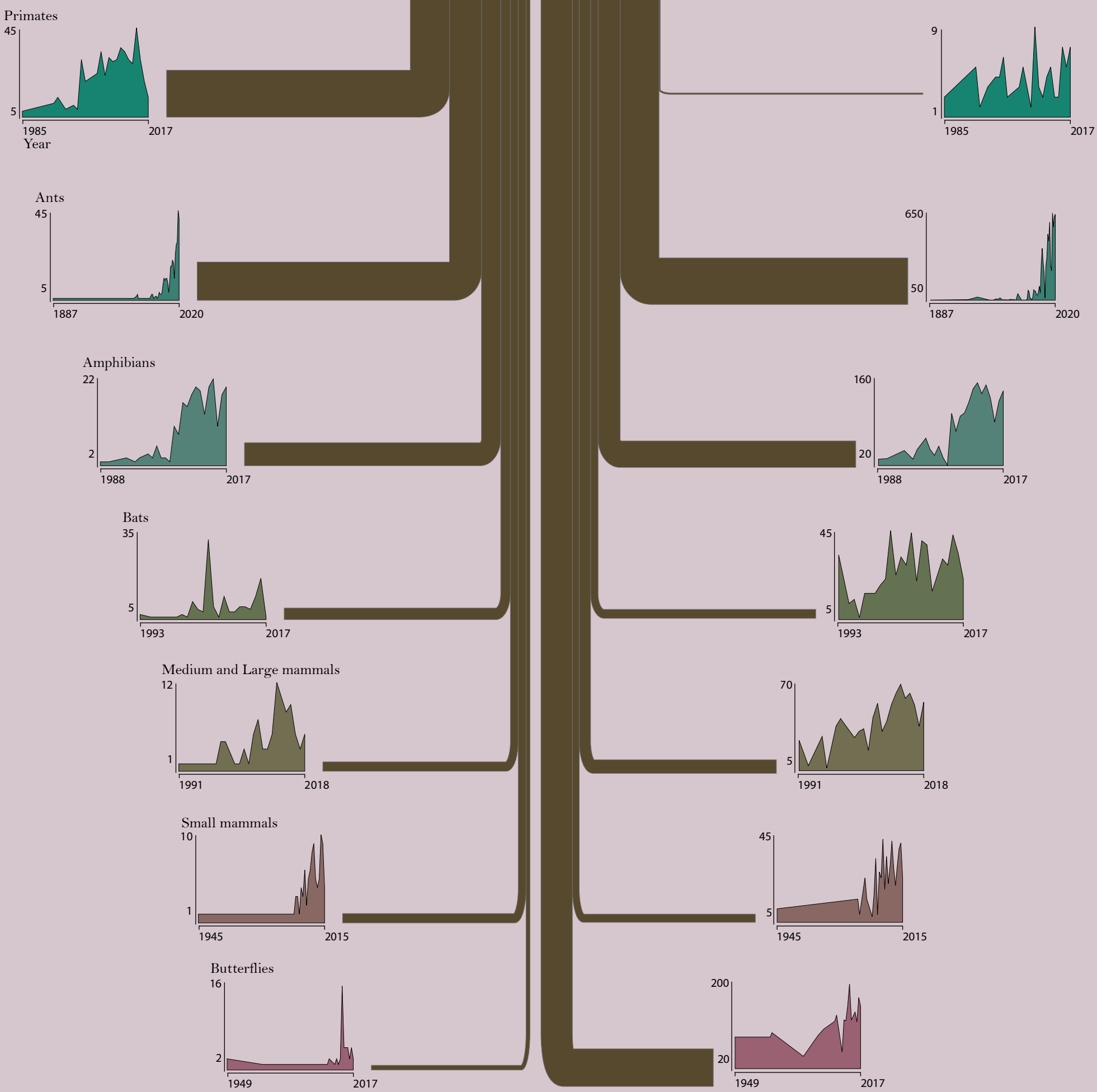
The almost lost and still unknown Brazilian Atlantic Forest



The first published paper about animals is dated 1987. Since then, the number of scientific publications has increased. Even so, more is needed to give us a good understanding of its species. The Atlantic Forest has lost over 85% of its original cover, but the publication number has not been consistent and balanced across taxa. Some groups of animals have received more attention than others.



The increase in the publication in scientific journals happened right after the millennium turn and ecological movements worldwide. Besides, there was a change in the government after this point, with more investments in education and opportunities for graduate students.



Version 2.

In the second version, I changed the canopy **bubble chart** to a **donut chart** showing the deforestation of the Atlantic forest. I also created vectors for each group I show in the **linear chart**. To make my tree more suitable, I included branches to make it more attractive and suitable. I used the branches to include boxes of information about the forest. I also added a *map* of the Atlantic forest covering the whole background of the infographic.

Atlantic Forest

The most threatened, almost lost, and still unknown Brazilian biome

The first published paper about animals is dated 1987. Since then, the number of scientific publications has increased. Even so, more is needed to give us a good understanding of its species

The forest harbors around 20,000 species of plants, with almost 450 tree species being found in just one hectare in some locations

Over 11,000 species of plants and animals are considered threatened today in the Atlantic Forest

Approximately 40 percent of its vascular plants and up to 60 percent of its vertebrates are endemic species, meaning they are found nowhere else in the world

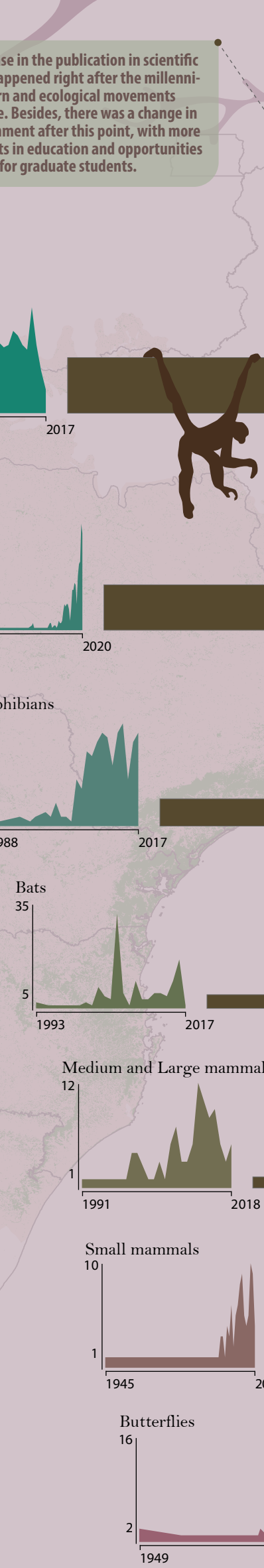
Nearly 250 species of amphibians, birds, and mammals have become extinct due to the result of human activity in the past 400 years

Around 85 % of the Atlantic Forest was lost. This deforestation continues at an annual rate of 0.5% and up to 2.9% in urban areas.

The 15% cover remaining comprise 150,000 km², which only 50% are inside protected areas

The increase in the publication in scientific journals happened right after the millennium turn and ecological movements worldwide. Besides, there was a change in the government after this point, with more investments in education and opportunities for graduate students.

The Atlantic Forest has lost over 85% of its original cover, but the publication number has not been consistent and balanced across taxa. Some groups of animals have received more attention than others



Version 3.

In version three, I removed the background *map* and put it in the middle of the **donut chart** with more information about other biomes in Brazil. Since the boxes around the texts in the branches looked weird, I removed them and left only the text. I also removed the background color and added some labels to make the charts more readable.

I changed the canopy **bubble chart** to a **donut chart** showing the deforestation of the Atlantic forest. I created vectors for each group I show in the **linear chart**. To make my tree more suitable, I included branches to make it more attractive and suitable. I used the branches to include boxes of information about the forest. I also added a map of the Atlantic forest covering the whole background of the infographic.

Atlantic Forest

The threatened, almost lost, and still unknown Brazilian biome

The first published paper about animals is dated 1987. Since then, the number of scientific publications has increased. Even so, more is needed to give us a good understanding of its species

Considered a Biodiversity hotspots because of its exceptional levels of plant endemism and serious levels of habitat loss

Around 85 % of the Atlantic Forest was lost. This deforestation continues at an annual rate of 0.5% and up to 2.9% in urban areas.

The 15% cover remaining comprise 150,000 km², which only 50% are inside protected areas

The increase in the publication in scientific journals happened right after the millennium turn and ecological movements worldwide. Besides, there was a change in the government after this point, with more investments in education and opportunities for graduate students.

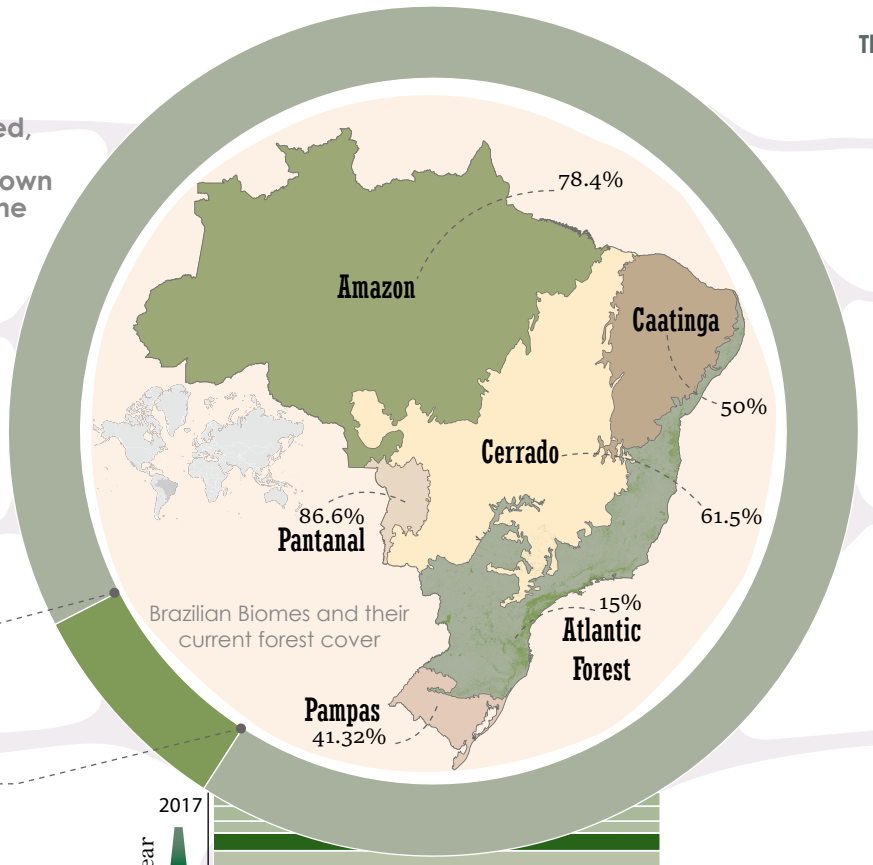
The forest harbors around 20,000 species of plants, with almost 450 tree species being found in just one hectare in some locations

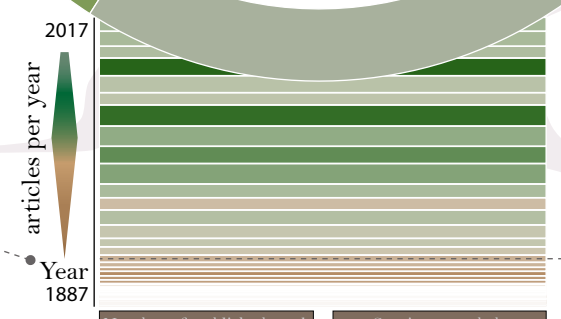
Approximately 40 percent of its vascular plants and up to 60 percent of its vertebrates are endemic species, meaning they are found nowhere else in the world

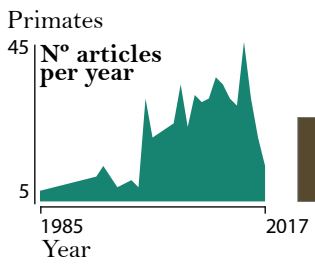
Over 11,000 species of plants and animals are considered threatened today in the Atlantic Forest

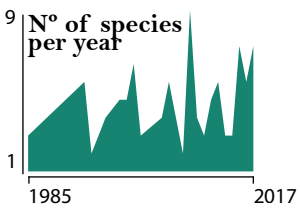
Nearly 250 species of amphibians, birds, and mammals have become extinct due to the result of human activity in the past 400 years

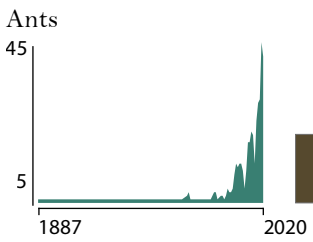
The Atlantic Forest has lost over 85% of its original cover, but the publication number has not been consistent and balanced across taxa. Some groups of animals have received more attention than others

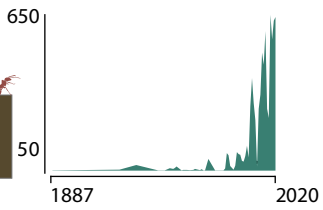


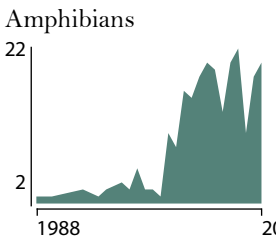


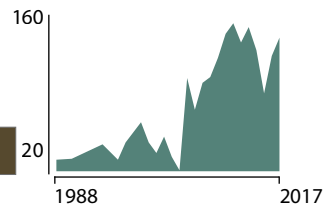


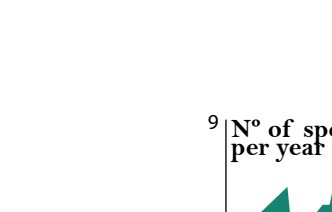


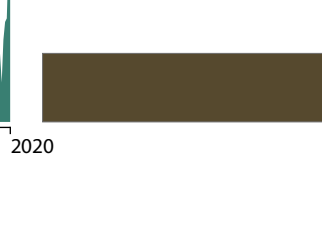


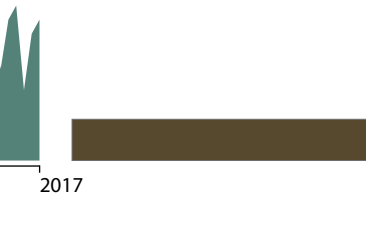


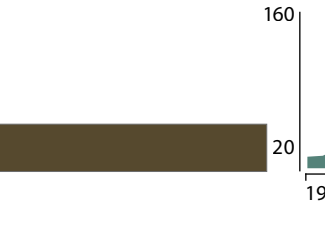












Version 4.

In this last version, I didn't make many significant changes. I just changed some colors of the animal vectors, played with the graph names to make them clearer, and changed the size of the graph title.

Atlantic Forest

The threatened, almost lost, and still unknown Brazilian biome. Due to the Atlantic Forest's vast diversity of endemic plants and animals as well as the fragmentation affecting these species, many groups and organizations are working to restore this unique ecosystem

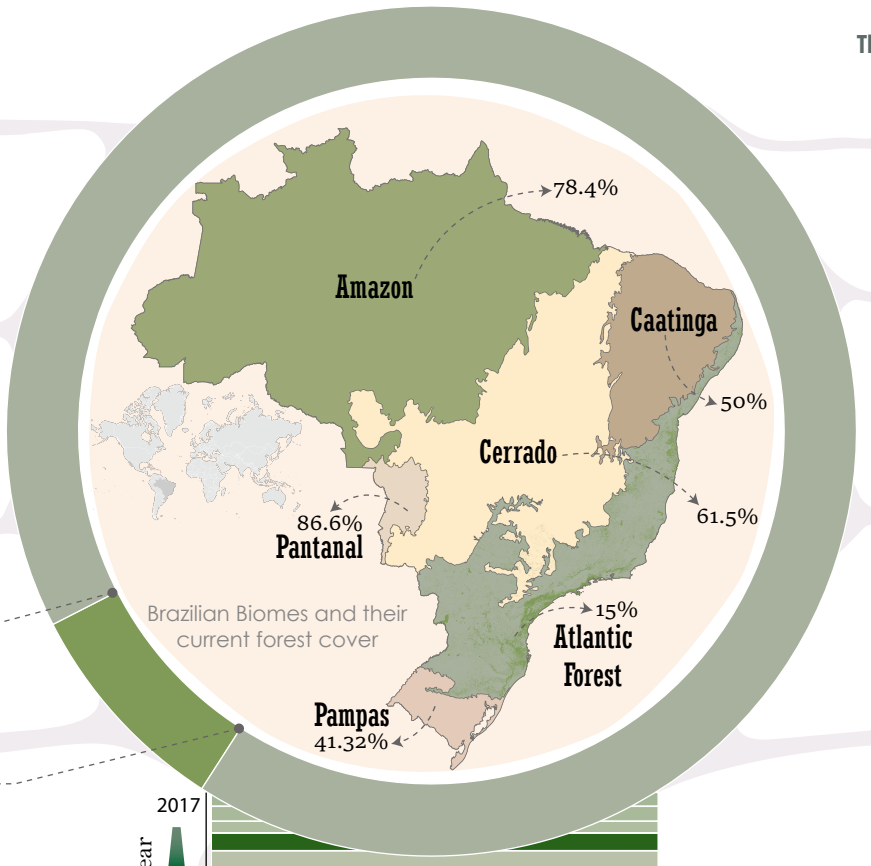
The forest harbors around 20,000 species of plants, with almost 450 tree species being found in just one hectare in some locations

Approximately 40 percent of its vascular plants and up to 60 percent of its vertebrates are endemic species**, meaning they are found nowhere else in the world**

The first published paper **about animals is dated 1987. Since then, the number of scientific publications has increased. Even so, more is needed to give us a good understanding of its species**

Considered a Biodiversity hotspots **because of its exceptional levels of plant endemism and serious levels of habitat loss**

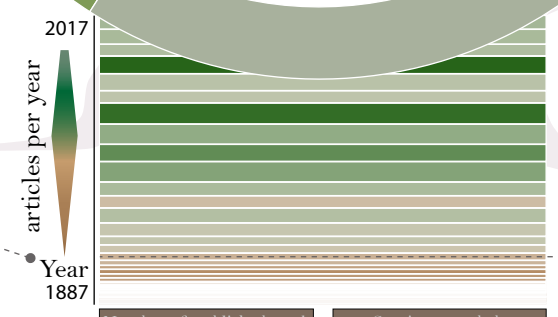
Over 11,000 species of plants and animals are considered threatened **today in the Atlantic Forest**



Around 85 % of the Atlantic Forest was lost. This deforestation **continues at an annual rate of 0.5% and up to 2.9% in urban areas.**

Nearly 250 species of amphibians, birds, and mammals have become extinct **due to the result of human activity in the past 400 years**

The 15% cover remaining comprise 150,000 km², which only 50% are inside protected areas



The increase in the publication in scientific journals happened right after the millennium turn and ecological movements **worldwide. Besides, there was a change in the government after this point, with more investments in education and opportunities for graduate students.**

The Atlantic Forest has lost **over 85% of its original cover, but the publication number has** not been consistent and balanced **across taxa. Some groups of animals have received more attention than others**

Design by: Yuri Souza
Source: https://esajourn als.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)1939-9170.AtlanticPapers

Personal thoughts and learning from this project

After watching the last speaker's talk, I tried to use data and art together. We always used to see human impacts in forests through charts such as **linear** or **bar charts** but did not integrate the complexity and relations of this information to the whole. Sometimes we might believe we are too familiar with these charts that we don't need to pay attention.

So, I made this infographic using different charts to create something more "alive" and informative. Something that someone could admire while learning. I also think that incorporating the **Sankey chart** with the **linear charts** make us wonder that charts displaying absolute information might hide important information that temporal charts can capture.

By looking at these charts, we can see something defined in the ecology field as a "*Cute species*", species that have an emotional impact on humans, such as primate species (the species more related to humans origin). Furthermore, primates were the group with more studies over time.

```
if(!require(tidyverse))install.packages("tidyverse", dependencies = TRUE)
```

packages

```
## Warning: package 'tidyverse' was built under R version 4.2.3

## Warning: package 'ggplot2' was built under R version 4.2.3

## Warning: package 'tibble' was built under R version 4.2.3

## Warning: package 'tidyr' was built under R version 4.2.3

## Warning: package 'readr' was built under R version 4.2.3

## Warning: package 'purrr' was built under R version 4.2.3

## Warning: package 'dplyr' was built under R version 4.2.3

## Warning: package 'forcats' was built under R version 4.2.3

## Warning: package 'lubridate' was built under R version 4.2.3
```

```
butterflies_paper <- read_csv("00_data/raw/ATLANTIC_BUTTERFLIES_references.csv") %>%
  dplyr::filter(Type == "Article") %>%
  dplyr::select(sites_ID, References) %>%
  dplyr::mutate(publised_year = gsub(".*?([0-9]+).*", "\\1", References)) %>%
  dplyr::select(!2)
```

Butterflies data

```
## Rows: 209 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): sites_ID, Authors, Type, References, Voucher specimens
## dbl (1): Ref_number
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
butterflies_paper_count <- butterflies_paper %>%
  dplyr::select(!1) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_p = n()) %>%
  dplyr::ungroup()

butterflies_spp <- read_csv2("00_data/raw/ATLANTIC_BUTTERFLIES_species.csv") %>%
```

```
  dplyr::select(sites_ID, Species) %>%
  dplyr::left_join(butterflies_paper) %>%
  dplyr::select(!1) %>%
  dplyr::group_by(publised_year) %>%
  distinct() %>%
  dplyr::summarize(total_s = n()) %>%
  dplyr::ungroup() %>%
  drop_na()
```

```
## i Using "','" as decimal and "'.'" as grouping mark. Use `read_delim()` for more control.
## New names:Rows: 7062 Columns: 16-- Column specification ----------------------------------------
## Delimiter: ";"
## chr (9): sites_ID, Class, Order, Family, Subfamily, Tribe, Genus, Species, S...
## lgl (7): ...10, ...11, ...12, ...13, ...14, ...15, ...16
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.Joining with `by = j
```

```
## Warning in dplyr::left_join(., butterflies_paper): Detected an unexpected many-to-many relationship b
## i Row 122 of `x` matches multiple rows in `y`.
## i Row 29 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
join_butterflies_spp_year <- dplyr::left_join(butterflies_spp, butterflies_paper_count) %>%
  write_csv("00_data/butterflies.csv")
```

```
## Joining with `by = join_by(publised_year)`
```

```
join_butterflies_spp_year
```

```
bats_paper <- read_csv("00_data/raw/ATLANTIC_BATS_Reference.csv") %>%
  dplyr::filter(Type == "Article") %>%
  dplyr::select(ID, References) %>%
  dplyr::mutate(publised_year = gsub(".*?([0-9]+).*", "\\1", References)) %>%
  dplyr::select(!2)
```

Bats data

```
## Rows: 205 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): ID, Authors, Type, References, Voucher_Specimens
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
bats_paper_count <- bats_paper %>%
  dplyr::select(!1) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_p = n()) %>%
  dplyr::ungroup()


bats_spp <- read_csv("00_data/raw/ATLANTIC_Bats_Capture.csv") %>%
  filter(Capture_number > 0) %>%
  dplyr::left_join(bats_paper) %>%
  dplyr::select(Species, publised_year) %>%
  drop_na() %>%
  dplyr::group_by(Species, publised_year) %>%
  distinct() %>%
  ungroup() %>%
  dplyr::select(!1) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_s = n())
```

```
## Rows: 21320 Columns: 20
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (16): ID, Duration, Season, Method, Strata_sampling, hours_per_night, nu...
## dbl  (4): Year_start, Year_finish, Nights, Capture_number
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Joining with `by = join_by(ID)`
```

```
#join_bats_spp_year <- dplyr::left_join(bats_spp, bats_paper_count) %>% write_csv("00_data/bats.csv")
```

```
primates_paper <- read_csv("00_data/raw/ATLANTIC-PR_References.csv") %>%
  dplyr::select(REF_ID, Referece) %>%
  dplyr::mutate(publised_year = as.numeric(gsub(".*?([0-9]+).*", "\\1", Referece))) %>%
  dplyr::filter(publised_year > 1000) %>%
  dplyr::mutate(publised_year = as.character(publised_year)) %>%
  dplyr::select(!2)
```

Primates data

```
## Rows: 586 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): REF_ID, PUB_TYPE, Referece
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
primates_paper_count <- primates_paper %>%
  dplyr::select(!1) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_p = n()) %>%
  dplyr::ungroup()


primates_spp <- read_csv2("00_data/raw/ATLANTIC-PR_Community.csv") %>%
  dplyr::left_join(primates_paper) %>%
  drop_na(publised_year) %>%
  dplyr::select(SPECIES, publised_year) %>%
  drop_na() %>%
  dplyr::group_by(SPECIES, publised_year) %>%
  distinct() %>%
  ungroup() %>%
  dplyr::select(!1) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_s = n())
```

```
## i Using "','" as decimal and "'.'" as grouping mark. Use `read_delim()` for more control.
## Rows: 2061 Columns: 31-- Column specification --------------------------------------------------------
## Delimiter: ";"
## chr (24): REF_ID, ORDEMBD, STUDY_AIM, SPECIES, SP_ORIGIN, SITE, AREA_HA, MUN...
## dbl  (4): ANNUAL_RAIN, ALTITUDE, COL_STRT_YR, COL_END_YR
## num  (3): LONGITUDE_X, LATITUDE_Y, ANNUAL_TEMP
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.Joining with `by = j
```

```r
#join_primates_spp_year <- dplyr::left_join(primates_spp, primates_paper_count) %>% write_csv("00_data/
```

```r
amphibians_paper <- read_csv("00_data/raw/ATLANTIC_AMPHIBIANS_references.csv") %>%
  dplyr::filter(reference_type == "a") %>%
  dplyr::select(reference_number, citation) %>%
  dplyr::mutate(publised_year = as.numeric(gsub(".*?([0-9]+).*", "\\1", citation)),
                reference_number = as.character(reference_number)) %>%
  dplyr::select(!2)
```

Amphibians data

```
## Rows: 389 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): reference_type, citation
## dbl (1): reference_number
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
amphibians_paper_count <- amphibians_paper %>%
  dplyr::select(!1) %>%
  dplyr::group_by(published_year) %>%
  dplyr::summarize(total_p = n()) %>%
  dplyr::ungroup()


amphibians_spp <- read_csv("00_data/raw/ATLANTIC_AMPHIBIANS_species.csv") %>%
  dplyr::select(c(1,8)) %>%
  drop_na() %>%
  dplyr::mutate(id = str_remove(id, "amp")) %>%
  dplyr::rename(reference_number = id,
                Species = valid_name) %>%
  dplyr::left_join(amphibians_paper) %>%
  dplyr::select(Species, published_year) %>%
  drop_na() %>%
  dplyr::group_by(Species, published_year) %>%
  distinct() %>%
  dplyr::select(!1) %>%
  dplyr::group_by(published_year) %>%
  dplyr::summarize(total_s = n())
```

```
## Rows: 17619 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (8): id, order, superfamily, family, subfamily, genus, species, valid_name
## dbl (2): individuals, endemic
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Joining with `by = join_by(reference_number)`
## Adding missing grouping variables: `Species`
```

```
join_amphibians_spp_year <- dplyr::left_join(amphibians_spp, amphibians_paper_count) %>% write_csv("00_
```

```
## Joining with `by = join_by(published_year)`
```

```
smallmammals_paper <- read_csv("00_data/raw/ATLANTIC_SM_Reference.csv") %>%
  dplyr::filter(reference_type == "Article") %>%
  dplyr::select(!3) %>%
  dplyr::mutate(published_year = as.character(published_year))
```

Small Mammals data

```
## Rows: 135 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): reference_type
## dbl (2): reference_number, published_year
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
smallmammals_paper_count <- smallmammals_paper %>%
  dplyr::select(!1) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_p = n()) %>%
  dplyr::ungroup()


smallmammals_spp <- read_csv("00_data/raw/ATLANTIC_SM_capture.csv") %>%
  dplyr::select(c(1,17)) %>%
  drop_na() %>%
  dplyr::left_join(smallmammals_paper) %>%
  dplyr::select(Actual_species_name, publised_year) %>%
  drop_na() %>%
  dplyr::group_by(Actual_species_name, publised_year) %>%
  distinct() %>%
  dplyr::select(!1) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_s = n())
```

```
## Rows: 2620 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (14): Reference_number, Month_start, Month_finish, Sampling_habitat, Sam...
## dbl  (7): reference_number, Year_start, Year_finish, Total_of_months, Effort...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Joining with `by = join_by(reference_number)`
## Adding missing grouping variables: `Actual_species_name`
```

```
join_smallmammals_spp_year <- dplyr::left_join(smallmammals_spp, smallmammals_paper_count) %>% write_csv
```

```
## Joining with `by = join_by(publised_year)`
```

```
birds_spp <- read_csv("00_data/raw/ATLANTIC_birds.csv") %>%
  dplyr::select(!1) %>%
  dplyr::group_by(Species, publised_year) %>%
  dplyr::distinct() %>%
  dplyr::ungroup() %>%
  dplyr::select(1) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_s = n()) %>%
  dplyr::ungroup()
```

Birds data

```
## Rows: 716 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Main_researcher, Species
## dbl (1): publised_year
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
birds_paper_count <- read_csv("00_data/raw/ATLANTIC_birds.csv") %>%
  dplyr::select(!3) %>%
  dplyr::group_by(Main_researcher, publised_year) %>%
  dplyr::distinct() %>%
  dplyr::summarize(total_p = n()) %>%
  dplyr::ungroup() %>%
  dplyr::select(publised_year) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_p = n()) %>%
  dplyr::ungroup()
```

```
## Rows: 716 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Main_researcher, Species
## dbl (1): publised_year
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## `summarise()` has grouped output by 'Main_researcher'. You can override using the `.groups` argument
```

```
join_birds_spp_year <- dplyr::left_join(birds_spp, birds_paper_count) %>% write_csv("00_data/birds.csv")
```

```
## Joining with `by = join_by(publised_year)`
```

```
mlmammals_spp <- read_csv("00_data/raw/ATLANTIC_ML_mammals.csv") %>%
  dplyr::select(!1) %>%
  dplyr::group_by(Species, publised_year) %>%
  dplyr::distinct() %>%
  dplyr::ungroup() %>%
  dplyr::select(1) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_s = n()) %>%
  dplyr::ungroup()
```

Medium and Large Mammals data

```
## Rows: 3412 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
```

```
## chr (2): Reference, Species
## dbl (1): publised_year
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
mlmammals_paper_count <- read_csv("00_data/raw/ATLANTIC_ML_mammals.csv") %>%
  dplyr::select(!3) %>%
  dplyr::group_by(Reference, publised_year) %>%
  dplyr::distinct() %>%
  dplyr::summarize(total_p = n()) %>%
  dplyr::ungroup() %>%
  dplyr::select(publised_year) %>%
  dplyr::group_by(publised_year) %>%
  dplyr::summarize(total_p = n()) %>%
  dplyr::ungroup()
```

```
## Rows: 3412 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Reference, Species
## dbl (1): publised_year
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## `summarise()` has grouped output by 'Reference'. You can override using the `.groups` argument.
```

```
join_mlmammals_spp_year <- dplyr::left_join(mlmammals_spp, mlmammals_paper_count) %>% write_csv("00_data
```

```
## Joining with `by = join_by(publised_year)`
```

```
all_data <- rbind(join_butterflies_spp_year, join_amphibians_spp_year, join_smallmammals_spp_year, join
  group_by(publised_year) %>%
  summarise(total_s = sum(total_s),
            total_p = sum(total_p)) %>%
  write_csv("00_data/total_data_summed.csv")
```

joining tables