Extinctions Over Time: A Data Analysis

Claire Parkinson

November 30, 2017

Introduction

The purpose of this post is to answer the following question: are we currently facing the sixth great extinction? To do this, we will also ask: are more species going extinct now than they have in the past? How does the extinction rate differ across different animal groups? Has the annual extinction rate increased? Is it accelerating? To answer all of these questions, we will use species data available online and various R programming and data analysis tools. We will give specific instructions and detailed comments to make all of the analysis in this post reproducible. An overview of the computational topics we will cover is below.

Computational Topics

- Accessing data from a RESTful API
- Error handling
- JSON data format
- Regular expressions
- Working with missing values
- Writing functions
- Reproducing statistics from online data

Data Preparation

Loading necessary packages

```
# Make sure you have all of these packages installed and load them in order to reproduce the same results

library("tidyverse")
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages --------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
library("httr")
library("jsonlite")
```

```
##
## Attaching package: 'jsonlite'
```

```
## The following object is masked from 'package:purrr':
##
##     flatten
```

```
library("reshape2")
```

```
## Warning: package 'reshape2' was built under R version 3.4.3
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

Extracting Data from online source

First, we will use the `httr` package to make a single API query against the following endpoint: `http://api.iucnredlist.org/index/species/Acaena-exigua.json` where all of the data on species extinction is stored publicly online.

```
resp <- httr::GET("http://api.iucnredlist.org/index/species/Acaena-exigua.json")

str(resp)
```

```
## List of 10
##  $ url        : chr "http://api.iucnredlist.org/index/species/Acaena-exigua.json"
##  $ status_code: int 200
##  $ headers    :List of 10
##   ..$ server          : chr "Cowboy"
##   ..$ date            : chr "Mon, 04 Dec 2017 03:47:29 GMT"
##   ..$ connection      : chr "keep-alive"
##   ..$ content-type    : chr "application/json; charset=utf-8"
##   ..$ etag            : chr "\"24dc8675574e40c835635bc220602e7f\""
##   ..$ cache-control   : chr "max-age=0, private, must-revalidate"
##   ..$ x-ua-compatible : chr "IE=Edge,chrome=1"
##   ..$ x-runtime       : chr "0.001865"
##   ..$ transfer-encoding: chr "chunked"
##   ..$ via             : chr "1.1 vegur"
##   ..- attr(*, "class")= chr [1:2] "insensitive" "list"
##  $ all_headers:List of 1
##   ..$ :List of 3
##   .. ..$ status : int 200
##   .. ..$ version: chr "HTTP/1.1"
##   .. ..$ headers:List of 10
##   .. .. ..$ server          : chr "Cowboy"
##   .. .. ..$ date            : chr "Mon, 04 Dec 2017 03:47:29 GMT"
##   .. .. ..$ connection      : chr "keep-alive"
##   .. .. ..$ content-type    : chr "application/json; charset=utf-8"
##   .. .. ..$ etag            : chr "\"24dc8675574e40c835635bc220602e7f\""
##   .. .. ..$ cache-control   : chr "max-age=0, private, must-revalidate"
##   .. .. ..$ x-ua-compatible : chr "IE=Edge,chrome=1"
##   .. .. ..$ x-runtime       : chr "0.001865"
##   .. .. ..$ transfer-encoding: chr "chunked"
##   .. .. ..$ via             : chr "1.1 vegur"
##   .. .. ..- attr(*, "class")= chr [1:2] "insensitive" "list"
##  $ cookies    :'data.frame': 0 obs. of  7 variables:
##   ..$ domain    : logi(0)
##   ..$ flag      : logi(0)
##   ..$ path      : logi(0)
##   ..$ secure    : logi(0)
##   ..$ expiration:Classes 'POSIXct', 'POSIXt'  num(0)
##   ..$ name      : logi(0)
##   ..$ value     : logi(0)
##  $ content    : raw [1:877] 5b 7b 22 73 ...
##  $ date       : POSIXct[1:1], format: "2017-12-04 03:47:29"
##  $ times      : Named num [1:6] 0 0.047 0.235 0.235 0.329 0.329
##   ..- attr(*, "names")= chr [1:6] "redirect" "namelookup" "connect" "pretransfer" ...
##  $ request    :List of 7
##   ..$ method   : chr "GET"
##   ..$ url      : chr "http://api.iucnredlist.org/index/species/Acaena-exigua.json"
##   ..$ headers  : Named chr "application/json, text/xml, application/xml, */*"
##   .. ..- attr(*, "names")= chr "Accept"
##   ..$ fields   : NULL
##   ..$ options  :List of 3
##   .. ..$ useragent: chr "libcurl/7.54.1 r-curl/2.8.1 httr/1.3.1"
##   .. ..$ cainfo   : chr "C:/PROGRA~1/R/R-34~1.1/etc/curl-ca-bundle.crt"
##   .. ..$ httpget  : logi TRUE
##   ..$ auth_token: NULL
##   ..$ output    : list()
##   .. ..- attr(*, "class")= chr [1:2] "write_memory" "write_function"
##   ..- attr(*, "class")= chr "request"
##  $ handle     :Class 'curl_handle' <externalptr>
##  - attr(*, "class")= chr "response"
```

```
# check what type of R object the data is

data <- content(resp)

class(data)
```

```
## [1] "list"
```

We can tell that the call was sucessful because the status code of 200 was returned in the output which indicates we successfully connected to the server. The return type object is a list, which is an R object. We could also return data as a data.frame using the fromJSON() function, demonstrated below.

```
# convert the data into a data frame

fromJSON(content(resp, as="text"))
```

```
##   scientific_name primary kingdom    phylum       class    order
## 1   Acaena exigua    TRUE PLANTAE TRACHEOPHYTA MAGNOLIOPSIDA ROSALES
##     family  genus species authority infra_name infra_rank infra_authority
## 1 ROSACEAE Acaena  exigua    A.Gray         NA         NA              NA
##   stock_name species_id main_common_name modified_year category criteria
## 1         NA      44072               NA          2003       EX       NA
##   assessmentid trend_id biome_marine biome_freshwater biome_terrestrial
## 1     10855072       NA        FALSE            FALSE              TRUE
##                                                                   taxonomicnotes
## 1 Held as "D" for now - need to confirm whether this was recently rediscovered. Have asked Vickie Caraway.
##                                                                        rationale
## 1 The last known individual of <em>A. exigua</em> was discovered in 1990s and has since died.
##                         assessor
## 1 Bruegmann, M.M. & Caraway, V.
##                                                                        evaluator
## 1 Maunder, M. (Plant Conservation Committee) & Dickinson, T. (Red List Programme Office)
```

Working with Regular Expressions

One of the entries in the response contains a field that may contain some information on when the species went extinct. We will identify the appropriate column and extract this information using a *regular expression*, as we have learned about in class.

example_string = "The last known individual of *A. exigua* was discovered in 1990s and has since died."

Regular Expression to match exactly 4 digits: `'\d{4}'`

```
# Use regular expression to extract date of extinction from example string

example_string = "The last known individual of <em>A. exigua</em> was discovered in 1990s and has since died."

stringr::str_extract(example_string, "\\d{4}")
```

```
## [1] "1990"
```

Calculating Extinction Rates: Putting it all together

First, to know what queries to make to the IUCN REST API, we need a list of extinct species names. This information can be downloaded from the IUCN website, but unfortunately this is not easily automated. Thus we'll download the data file using a public copy already prepared on github.

```
extinct = read_csv("https://espm-157.github.io/extinction-module/extinct.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   `Species ID` = col_integer(),
##   `Red List criteria version` = col_double(),
##   `Year assessed` = col_integer()
## )
```

```
## See spec(...) for full column specifications.
```

```
extinct
```

```
## # A tibble: 834 x 23
##    `Species ID` Kingdom    Phylum        Class         Order
##           <int> <chr>      <chr>         <chr>         <chr>
## 1         44072 PLANTAE  TRACHEOPHYTA MAGNOLIOPSIDA        ROSALES
## 2        195373 PLANTAE  TRACHEOPHYTA MAGNOLIOPSIDA    EUPHORBIALES
## 3         37854 PLANTAE  TRACHEOPHYTA MAGNOLIOPSIDA    EUPHORBIALES
## 4        199821 PLANTAE  TRACHEOPHYTA MAGNOLIOPSIDA    EUPHORBIALES
## 5            82 ANIMALIA   ARTHROPODA        INSECTA   EPHEMEROPTERA
## 6           167 ANIMALIA     MOLLUSCA     GASTROPODA STYLOMMATOPHORA
## 7           170 ANIMALIA     MOLLUSCA     GASTROPODA STYLOMMATOPHORA
## 8           173 ANIMALIA     MOLLUSCA     GASTROPODA STYLOMMATOPHORA
## 9           174 ANIMALIA     MOLLUSCA     GASTROPODA STYLOMMATOPHORA
## 10          179 ANIMALIA     MOLLUSCA     GASTROPODA STYLOMMATOPHORA
## # ... with 824 more rows, and 18 more variables: Family <chr>,
## #   Genus <chr>, Species <chr>, Authority <chr>, `Infraspecific
## #   rank` <chr>, `Infraspecific name` <chr>, `Infraspecific
## #   authority` <chr>, `Stock/subpopulation` <chr>, Synonyms <chr>, `Common
## #   names (Eng)` <chr>, `Common names (Fre)` <chr>, `Common names
## #   (Spa)` <chr>, `Red List status` <chr>, `Red List criteria` <chr>, `Red
## #   List criteria version` <dbl>, `Year assessed` <int>, `Population
## #   trend` <chr>, Petitioned <chr>
```

We will now write a function to extract the rationale for the extinction for all extinct species in the publically available data set online.

```r
get_rationale <- function(Genus, Species){
  resp <- GET(paste0("http://api.iucnredlist.org/index/species/", Genus, "-", Species, ".json")) # pastes link wit
h each Genus/Specie combination to get all data
  if(httr::status_code(resp) != 200){ # status code of not 200 indictes connection to server failed, so data on th
at specie is missing and denoted with NA
    return("NA")
  }
  data <- content(resp)
  out <- data[[1]]$rationale
  if(is.null(out))  # returns NA if 'rationale' information is missing for that specie
    return("NA")

  out
}
```

We will now test our function on a subset of the data before attempting the full data set. We will use our `dplyr` pipe syntax to iterate over our function.

```r
df_small <-
extinct[1:5,] %>% # subsets the first 5 rows of data
  rowwise() %>% # allows you to use mutate() with a list variable
  mutate(rationale = get_rationale(Genus, Species)) # adds new column with result from get_rationale() function us
ing each row's Genus and Species

df_small
```

```
## Source: local data frame [5 x 24]
## Groups: <by row>
##
## # A tibble: 5 x 24
##    `Species ID`  Kingdom        Phylum         Class        Order
##           <int>    <chr>         <chr>         <chr>        <chr>
## 1       44072  PLANTAE TRACHEOPHYTA MAGNOLIOPSIDA      ROSALES
## 2      195373  PLANTAE TRACHEOPHYTA MAGNOLIOPSIDA  EUPHORBIALES
## 3       37854  PLANTAE TRACHEOPHYTA MAGNOLIOPSIDA  EUPHORBIALES
## 4      199821  PLANTAE TRACHEOPHYTA MAGNOLIOPSIDA  EUPHORBIALES
## 5          82 ANIMALIA    ARTHROPODA       INSECTA EPHEMEROPTERA
## # ... with 19 more variables: Family <chr>, Genus <chr>, Species <chr>,
## #   Authority <chr>, `Infraspecific rank` <chr>, `Infraspecific
## #   name` <chr>, `Infraspecific authority` <chr>,
## #   `Stock/subpopulation` <chr>, Synonyms <chr>, `Common names
## #   (Eng)` <chr>, `Common names (Fre)` <chr>, `Common names (Spa)` <chr>,
## #   `Red List status` <chr>, `Red List criteria` <chr>, `Red List criteria
## #   version` <dbl>, `Year assessed` <int>, `Population trend` <chr>,
## #   Petitioned <chr>, rationale <chr>
```

We can see that our function get_rationale() was successful on the test set since we now have a new column "rationale" that has the information about when that species went extinct.

Now we will create a function that can extract the date from the rationale, and include this function in our data analysis pipeline.

```r
# This function extracts the date of extinction using regular expressions from our new 'rationale' column

last_seen <- function(rationale){
  date <- stringr::str_extract(rationale, "\\d{4}") # extracts 4 digit date from string
  as.numeric(date) # converts date from string to numeric variable
}
```

With these two new functions, we will now use them on the entire data set to get the extinction dates for each specie, in a new column called 'last_seen.'

```r
df <-
extinct %>%
  rowwise() %>%
  mutate(rationale = get_rationale(Genus, Species),
         last_seen = last_seen(rationale))
head(df)
```

```
## # A tibble: 6 x 25
##   `Species ID` Kingdom      Phylum       Class             Order
##          <int> <chr>        <chr>        <chr>             <chr>
## 1        44072 PLANTAE  TRACHEOPHYTA MAGNOLIOPSIDA          ROSALES
## 2       195373 PLANTAE  TRACHEOPHYTA MAGNOLIOPSIDA     EUPHORBIALES
## 3        37854 PLANTAE  TRACHEOPHYTA MAGNOLIOPSIDA     EUPHORBIALES
## 4       199821 PLANTAE  TRACHEOPHYTA MAGNOLIOPSIDA     EUPHORBIALES
## 5           82 ANIMALIA   ARTHROPODA       INSECTA     EPHEMEROPTERA
## 6          167 ANIMALIA     MOLLUSCA     GASTROPODA STYLOMMATOPHORA
## # ... with 20 more variables: Family <chr>, Genus <chr>, Species <chr>,
## #   Authority <chr>, `Infraspecific rank` <chr>, `Infraspecific
## #   name` <chr>, `Infraspecific authority` <chr>,
## #   `Stock/subpopulation` <chr>, Synonyms <chr>, `Common names
## #   (Eng)` <chr>, `Common names (Fre)` <chr>, `Common names (Spa)` <chr>,
## #   `Red List status` <chr>, `Red List criteria` <chr>, `Red List criteria
## #   version` <dbl>, `Year assessed` <int>, `Population trend` <chr>,
## #   Petitioned <chr>, rationale <chr>, last_seen <dbl>
```
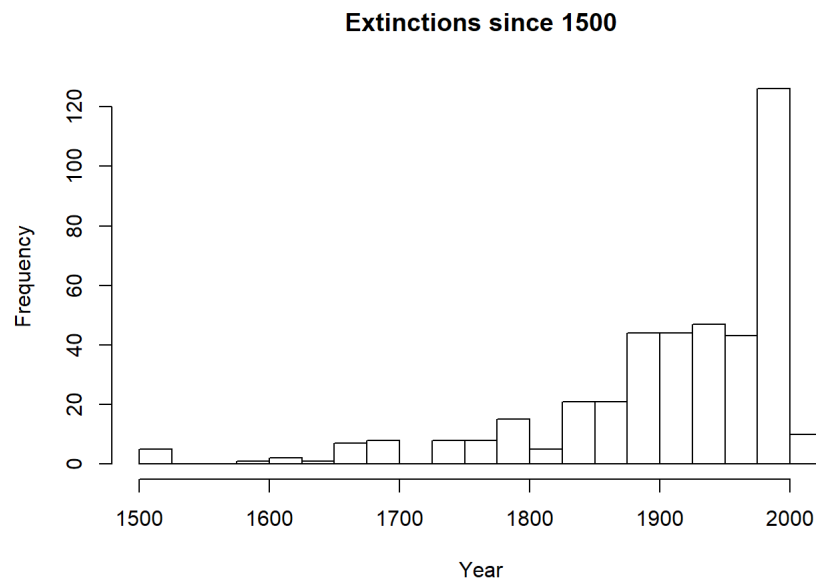
Histogram of Extinction Dates

We can visualize the tempo of extinctions by plotting extinctions since 1500 in 25-year interval bins in a histogram.

```
df <- df %>%
  filter(last_seen > 1500)

bins <- seq(1500, 2025, 25)

hist(na.omit(df$last_seen), breaks = bins, main = 'Extinctions since 1500', xlab = 'Year')
```


Extinctions since 1500

From this histogram, we can see that the number of extinctions per year has increased steadily over time, and more quickly since 1900. There has also been an extremely high jump in extinctions in the late 1900's. The reason that the last bin is so small is likely since the most recent data point is 2015 and the bin ends in 2025, 8 years in the future where we don't data for yet.

Extinctions by type of species

To see how extinctions differ across different types of species, we will compute the number of extinctions from 1500 - 1900 and from 1900 to to present-day for each of the following taxonomic groups:

- Vertebrates
- Mammals
- Birds
- Fish
- Amphibians
- Reptiles
- Insects
- Plants

Computing extinctions from 1500-1900

```r
# filters data frame to only include data on extinctions between 1500 and 1900
to_1900 <- df %>%
  filter(last_seen > 1500) %>%
  filter(last_seen < 1900)

paste("Number of Extinctions by taxonomic groups from 1500 to 1900")
```

```
## [1] "Number of Extinctions by taxonomic groups from 1500 to 1900"
```

```r
vert <- to_1900 %>%
  filter(Phylum == 'CHORDATA')
paste("Number of extinct vertebrates:" , nrow(vert))
```

```
## [1] "Number of extinct vertebrates: 114"
```

```r
mamm <- to_1900 %>%
  filter(Class == 'MAMMALIA')
paste("Number of extinct mammals:", nrow(mamm))
```

```
## [1] "Number of extinct mammals: 22"
```

```r
birds <- to_1900 %>%
  filter(Class == 'AVES')
paste("Number of extinct birds:", nrow(birds))
```

```
## [1] "Number of extinct birds: 83"
```

```r
fish <- to_1900 %>%
  filter(Class == 'ACTINOPTERYGII')
paste("Number of extinct fish:", nrow(fish))
```

```
## [1] "Number of extinct fish: 2"
```

```r
amphib <- to_1900 %>%
  filter(Class == 'AMPHIBIA')
paste("Number of extinct amphibians:", nrow(amphib))
```

```
## [1] "Number of extinct amphibians: 0"
```

```r
reptile <- to_1900 %>%
  filter(Class == 'REPTILIA')
paste("Number of extinct reptiles:", nrow(reptile))
```

```
## [1] "Number of extinct reptiles: 7"
```

```r
insect <- to_1900 %>%
  filter(Class == 'INSECTA')
paste("Number of extinct insects:", nrow(insect))
```

```
## [1] "Number of extinct insects: 1"
```

```r
plant <- to_1900 %>%
  filter(Kingdom == 'PLANTAE')
paste("Number of extinct plants:", nrow(plant))
```

```
## [1] "Number of extinct plants: 8"
```

```r
paste("Total number of extinctions:", nrow(to_1900))
```

```
## [1] "Total number of extinctions: 143"
```

Computing extinctions since 1900

```r
# filters data frame to only include data on extinctions since 1900

since_1900 <- df %>%
  filter(last_seen > 1900)

paste("Number of Extinctions by taxonomic groups since 1900")
```

```
## [1] "Number of Extinctions by taxonomic groups since 1900"
```

```
vert <- since_1900 %>%
  filter(Phylum == 'CHORDATA')
paste("Number of extinct vertebrates:" , nrow(vert))
```

```
## [1] "Number of extinct vertebrates: 120"
```

```
mamm <- since_1900 %>%
  filter(Class == 'MAMMALIA')
paste("Number of extinct mammals:", nrow(mamm))
```

```
## [1] "Number of extinct mammals: 29"
```

```
birds <- since_1900 %>%
  filter(Class == 'AVES')
paste("Number of extinct birds:", nrow(birds))
```

```
## [1] "Number of extinct birds: 44"
```

```
fish <- since_1900 %>%
  filter(Class == 'ACTINOPTERYGII')
paste("Number of extinct fish:", nrow(fish))
```

```
## [1] "Number of extinct fish: 33"
```

```
amphib <- since_1900 %>%
  filter(Class == 'AMPHIBIA')
paste("Number of extinct amphibians:", nrow(amphib))
```

```
## [1] "Number of extinct amphibians: 10"
```

```
reptile <- since_1900 %>%
  filter(Class == 'REPTILIA')
paste("Number of extinct reptiles:", nrow(reptile))
```

```
## [1] "Number of extinct reptiles: 4"
```

```
insect <- since_1900 %>%
  filter(Class == 'INSECTA')
paste("Number of extinct insects:", nrow(insect))
```

```
## [1] "Number of extinct insects: 9"
```

```
plant <- since_1900 %>%
  filter(Kingdom == 'PLANTAE')
paste("Number of extinct plants:", nrow(plant))
```

```
## [1] "Number of extinct plants: 28"
```

```
paste("Total number of extinctions:", nrow(since_1900))
```

```
## [1] "Total number of extinctions: 270"
```

To test the accuracy of our estimates, we can compare them to Table 1 of Ceballos et al (2015).

Making a barchart

To better visualize how the number of extinctions has changed in each time period by taxonomic group, we can make a barchart.

```
# First we will make a data frame with the information we computed above

groups <- c('Vertebrates', 'Mammals', 'Birds', 'Fish', 'Amphibians', 'Reptiles', 'Insects', 'Plants')

extinctions1 <- c(114, 22, 83, 2, 0, 7, 1, 8) # extinctions from 1500 to 1900

extinctions2 <- c(120, 29, 44, 33, 10, 4, 9, 28) # extinctions from 1900 to today

extinctions <- data.frame(groups, extinctions1, extinctions2)

extinctions
```

```
##        groups extinctions1 extinctions2
## 1 Vertebrates          114          120
## 2     Mammals           22           29
## 3       Birds           83           44
## 4        Fish            2           33
## 5  Amphibians            0           10
## 6    Reptiles            7            4
## 7     Insects            1            9
## 8      Plants            8           28
```

Now that we have our data in a data frame, we can put it into a barchart using ggplot. In order to have 2 bars side-by-side for each taxonomic group, we will use the function melt() to convert our data frame into a "molten" data frame that ggplot can use to make multiple bars per x-axis object.

```
ext <- melt(extinctions)
```

```
## Using groups as id variables
```

```
names(ext)[3] <- "extinctions" # rename value column

ext
```

```
##           groups      variable extinctions
## 1  Vertebrates extinctions1          114
## 2      Mammals extinctions1           22
## 3        Birds extinctions1           83
## 4         Fish extinctions1            2
## 5   Amphibians extinctions1            0
## 6     Reptiles extinctions1            7
## 7      Insects extinctions1            1
## 8       Plants extinctions1            8
## 9  Vertebrates extinctions2          120
## 10     Mammals extinctions2           29
## 11       Birds extinctions2           44
## 12        Fish extinctions2           33
## 13  Amphibians extinctions2           10
## 14    Reptiles extinctions2            4
## 15     Insects extinctions2            9
## 16      Plants extinctions2           28
```

Now, we can use ggplot to make our bar chart comparing the number of extinctions per taxonomic group in the two time periods. We will also create a legend and make the x-axis labels vertical so they are easier to read.

```
# we are now ready to make our barchart
# we will go through and explain each step through comments

ggplot(ext, aes(x = groups, y = extinctions, fill = variable), xlab="Age Group") + # plot data from melted table '
ext'

geom_bar(stat="identity", width = .7, position = "dodge") + # add bars

  labs(x = "Taxonomic Group", y = "Number of extinctions", title = "Extinctions by animal types over time") + # re
label axes and add title

  scale_fill_discrete(name="Time\nPeriod", labels = c("1500 - 1900", "1900 - today")) + # relabel legend

  theme(text = element_text(size=10), # set font size
        axis.text.x = element_text(angle=90, hjust=1)) # make x-axis labels be written vertically
```
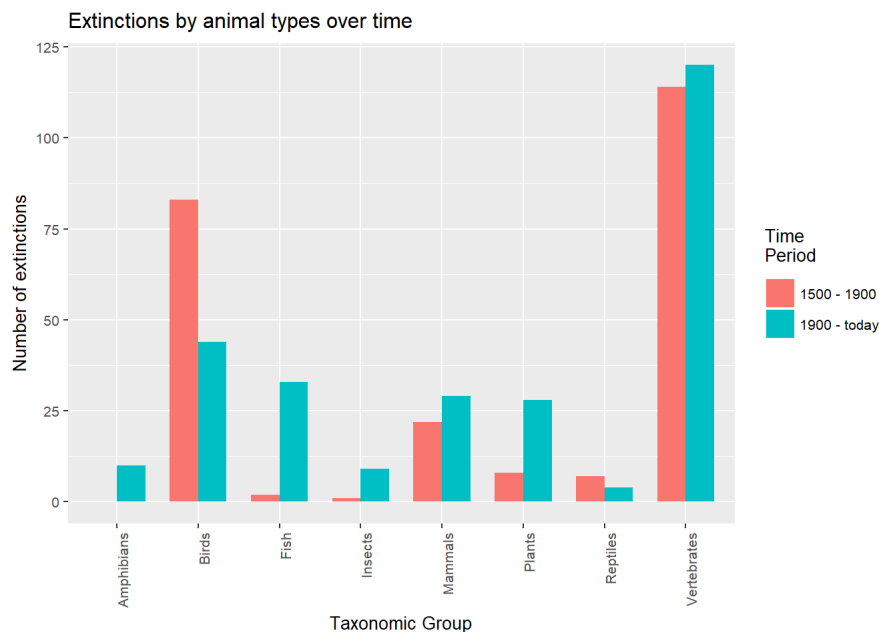
Extinctions by animal types over time

Looking at this graph, we can see that many more fish and plant species have gone extinct since 1900 then between 1500 and 1900. More amphibian and insect species have also gone extinct. Many more species of birds, on the other hand, went extinct before 1900 and about the same number of vertebrates have gone extinct. However, since the sizes of the time periods are vastly different (400 vs 115 years), these numbers will very likely soon surpass the previous time period.

Overall extinction rate

From our data, we can also compute the overall extinction rate for each time period and compare them.

```
paste("The average number of extinctions per year from 1500-1900 was:", round(143/400, 2))
```

```
## [1] "The average number of extinctions per year from 1500-1900 was: 0.36"
```

```
paste("The average number of extinctions per year from 1900 to today is:", round(270/115, 2))
```

```
## [1] "The average number of extinctions per year from 1900 to today is: 2.35"
```

```
paste0("The annual extinction rate has increased by: ", round((2.35-0.36)/0.36, 2) * 100,"%")
```

```
## [1] "The annual extinction rate has increased by: 553%"
```

Conclusion

In this post, we have attempted to answer the question: are we currently experiencing the sixth great extinction in Earth's history? We pulled extensive and public data online (in JSON format) from a RESTful API about species and their extinction dates. We analyzed this data using regular expressions, various dyplr functions, and by creating our own functions. We visualized this data in many ways from lists and data frames to histograms and barcharts. Overall, our data analysis and R programming has revealed to us a frightening reality: more species have gone extinct in the past 100 years than the previous 400 years. Fish and plant species especially have suffered since 1900. The average number of extinctions per year between 1500 and 1900 and 1900 to today has increased by a whopping 553%. We could very well be facing the sixth great extinction. This conclusion requires us to ask of ourselves yet another question: what are we going to do about it?

References

Background information: - Introductory Video - Ceballos et al (2015)

On working with regular expressions: - Self-guided Tutorial - Cheetsheet - stringr RegEx Vignette

On using ggplot: - Graphs with ggplot2 - Bar charts