

HW3-drafting-viz

AUTHOR
Madi Calbert

PUBLISHED
February 10, 2025

Questions

Question 1

Which option do you plan to pursue? It's okay if this has changed since HW 1. Infographic

- Option 1 (infographic)!

Question 2

Restate your question(s). Has this changed at all since HW #1? If yes, how so?

- How has the international trade of the top threatened and endangered big cats (*Felidae* family) evolved over the years?
 - most common species of threatened and endangered big cats?
 - number of reported cats across time?
 - tops exporting/importing countries?
- Yes, my question has somewhat changed. I am now focusing on Appendix I cats (threatened and endangered species). I also realized I needed to look at the gross exports of species rather than a comparison between import and exports for my total counts across species and time. So I updated my data set to account for that.

Question 3

Explain which variables from your data set(s) you will use to answer your question(s), and how.

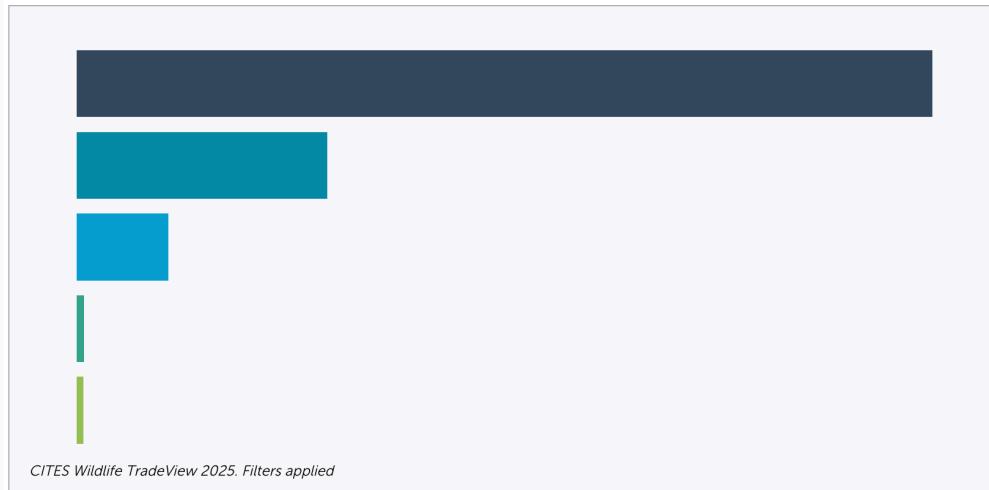
- Most reported big cats: I grouped by taxon and summed the counts of the number of specimens. I then isolated the top 5 species and plotted the number of specimens for each species.
- Reported cats across time: I grouped by year to visualize any trends of reported big cat trade over the years.
- Top exporters: I filtered my data down to just between 2020 and 2024. Then I filtered for countries that exported over 50 animals. Here I am using exporter, importer, and the importer reported quantity.

Question 4.

In HW #2, you created some exploratory data viz to better understand your data. You may already have some ideas of how you plan to formally visualize your data, but it's incredibly helpful to look at visualizations by other creators for inspiration. Find at least two data visualizations that you could (potentially) borrow / adapt pieces from. Link to them or download and embed them into your .qmd file, and explain which elements you might borrow (e.g. the graphic form, legend design, layout, etc.).

This website [CITES Wildlife Tradeview](#) has several visualizations of the CITES database that I referenced for my own infographic.

Top Trade Terms


[Download](#)
[Full list](#)

1. **Specimens** (9,411)

2. **Trophies** (2,758)

3. **Live** (1,011)

4. **Skulls** (80)

5. **Skins** (75)

Top Exporting Countries


[Download](#)
[Full list](#)

1. **Mozambique** (4,926)

2. **Namibia** (3,804)

3. **Lao People's Democratic Republic** (622)

4. **Zambia** (598)

5. **South Africa** (597)

Top Exporting Countries

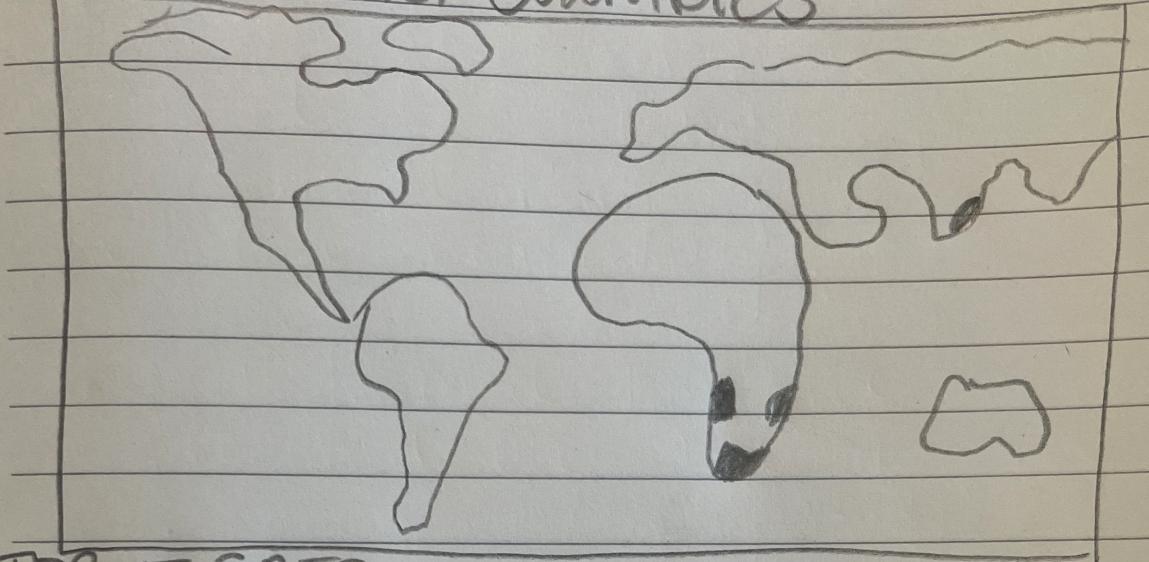
Question 5.

Hand-draw your anticipated visualizations, then take a photo of your drawing(s) and embed it in your rendered .qmd file – note that these are not exploratory visualizations, but rather your plan for your final visualizations that you will eventually polish and submit with HW #4. You should have: a sketch of your infographic (which should include at least three component visualizations) if you are pursuing option 1

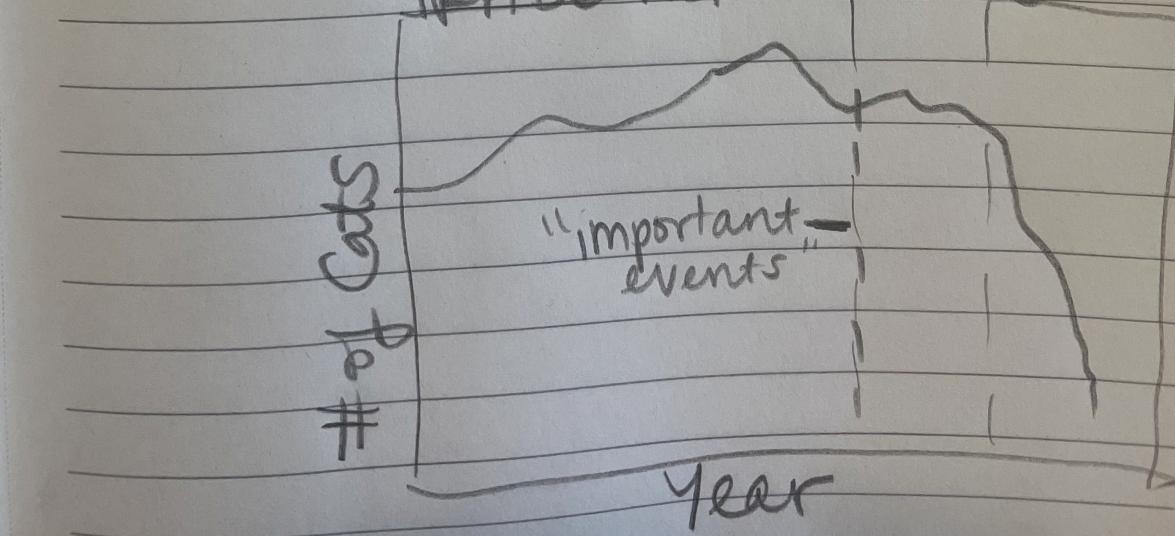
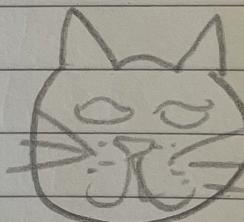
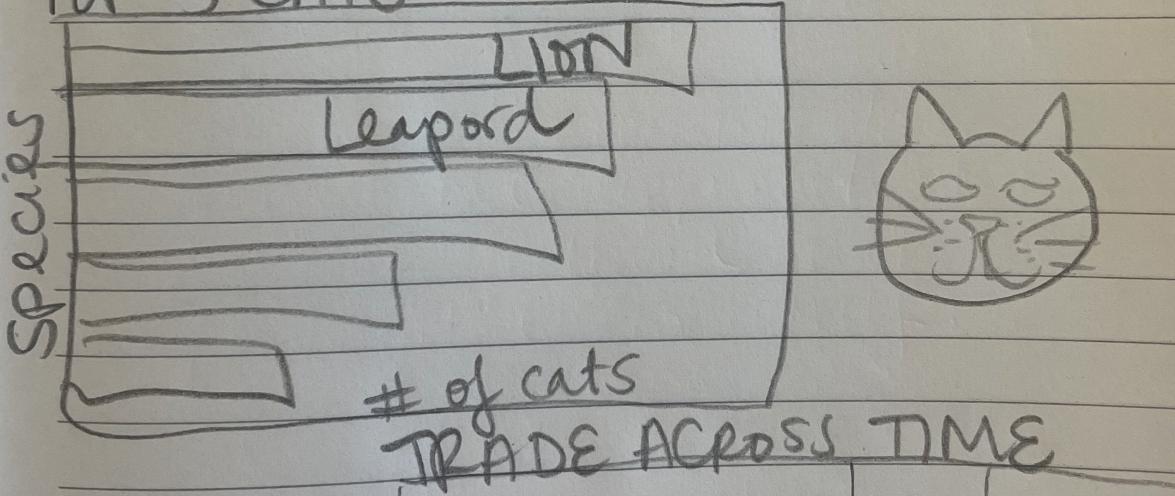
BIG CATS

TOP EXPORTING COUNTRIES

CITES



TOP 5 CATS



Mock up drawing of Final Infographic

Question 6

Mock up all of your hand drawn visualizations using code. We understand that you will continue to iterate on these into HW #4 (particularly after receiving feedback), but by the end of HW #3, you should:

- have your data plotted (if you're experimenting with a graphic form(s) that was not explicitly covered in class, we understand that this may take some more time to build; you should have as much put together as possible)
- use appropriate strategies to highlight / focus attention on a clear message
- include appropriate text such as titles, captions, axis labels
- experiment with colors and typefaces / fonts
- create a presentable / aesthetically-pleasing theme (e.g. (re)move gridlines / legends as appropriate, adjust font sizes, etc.)

Set up

```
# load libraries
library(tidyverse)
library(here)
library(janitor)
library(scales)
library(networkD3)
library(dplyr)

# load data - comparative (imports / exports) data
cites <- read.csv(here("data/cites_felidae_1975_2024.csv")) %>%
  clean_names()

# load data - gross exports data
gross_exports <- read.csv(here("data/gross_exports_1975_2025.csv")) %>%
  clean_names()

# clean up data
gross_exports_clean <- gross_exports %>%
  rename_with(~ str_remove(.x, "^\$")) %>%
  pivot_longer(
    cols = -c("app", "taxon", "term", "unit", "country"),
    names_to = "year",
    values_to = "count"
  ) %>%
  mutate(year = as.integer(year)) %>%
  drop_na() %>%
  filter(unit == "Number of specimens" | unit == "") %>%
  filter(app == "I") %>%
  filter(term %in% c("trophies", "live", "specimens"))

gross_exports_long <- gross_exports_clean %>%
```

```
mutate(count = as.integer(count)) %>%
uncount(count)
```

Top Exports

"Appendix I" refers to a list of species considered to be most critically endangered and threatened with extinction

```
# Add in the species common names
common_names <- data.frame(
  taxon = c(
    "Panthera tigris",
    "Acinonyx jubatus",
    "Panthera pardus",
    "Panthera onca",
    "Lynx pardinus"),
  common_name = c(
    "Tiger",
    "Cheetah",
    "Leopard",
    "Jaguar",
    "Iberian lynx"))

# Count occurrences of each taxon and get the top 5
top_5 <- gross_exports_long %>%
  count(taxon, sort = TRUE) %>%
  slice_max(n, n = 5)

top_5 <- top_5 %>%
  left_join(common_names, by = "taxon")

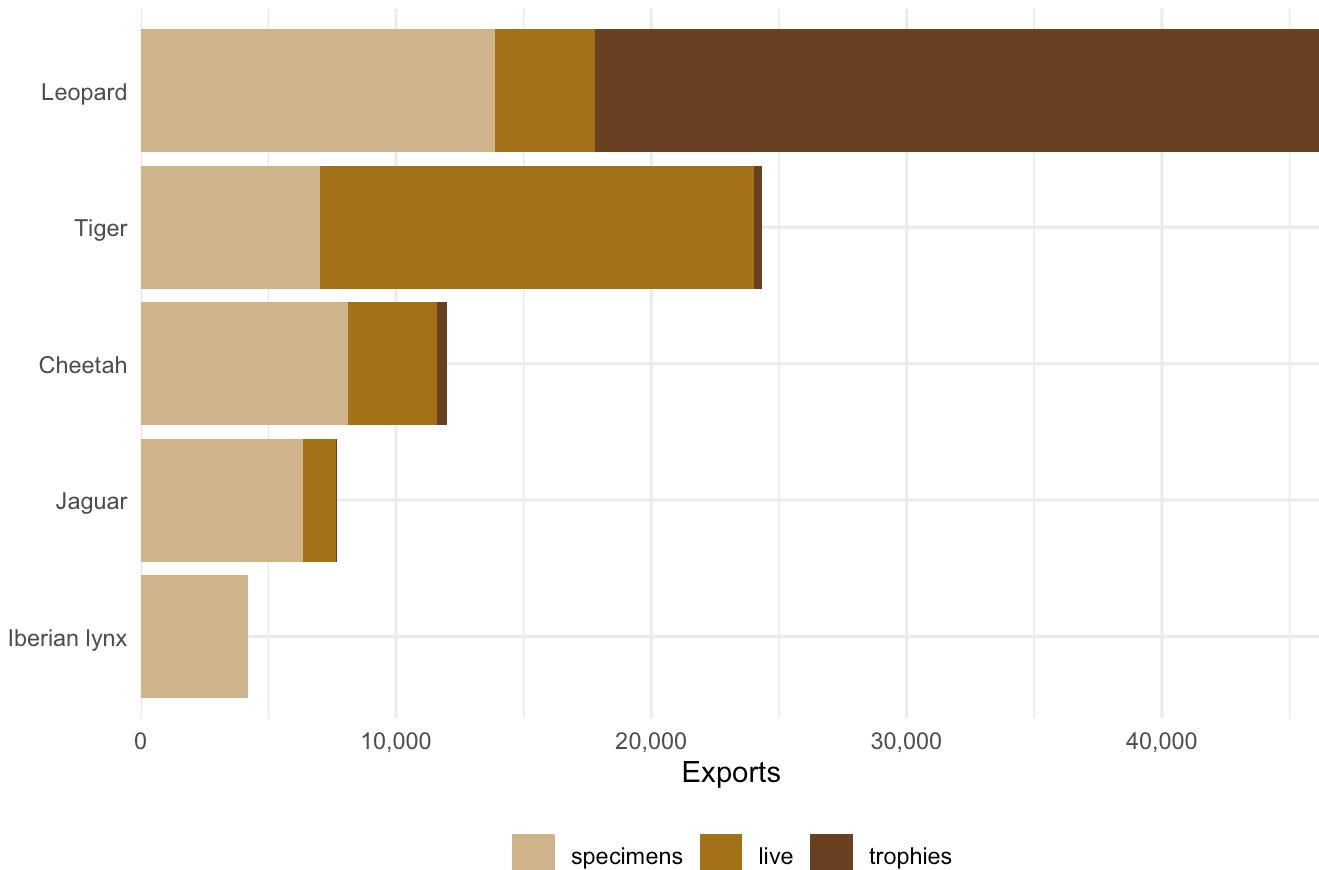
# Make a data frame that is the top 5 species and the top 5 terms
top_5_taxa_terms <- gross_exports_long %>%
  filter(taxon %in% top_5$taxon) %>%
  count(taxon, term, sort = TRUE) %>%
  left_join(common_names, by = "taxon") %>%
  group_by(common_name) %>%
  mutate(total_count = sum(n)) %>%
  ungroup()

my_pal <- scale_fill_manual(values = c(
  "specimens" = "#D2B48C",
  "live" = "#A6761D",
  "trophies" = "#6B4226"))

top_5_taxa_terms %>%
  ggplot(aes(x = fct_reorder(common_name, total_count),
             y = n,
             fill = fct_reorder(term, n))) +
```

```
geom_col() +
coord_flip() +
labs(title = "International Trade of Threatened & Endangered Big Cats",
x = NULL,
y = "Exports",
fill = NULL) +
theme_minimal() +
scale_y_continuous(expand = expansion(mult = c(0,0)),
labels = scales::comma) +
theme(legend.position = "bottom") +
my_pal +
guides(fill = guide_legend(reverse = TRUE))
```

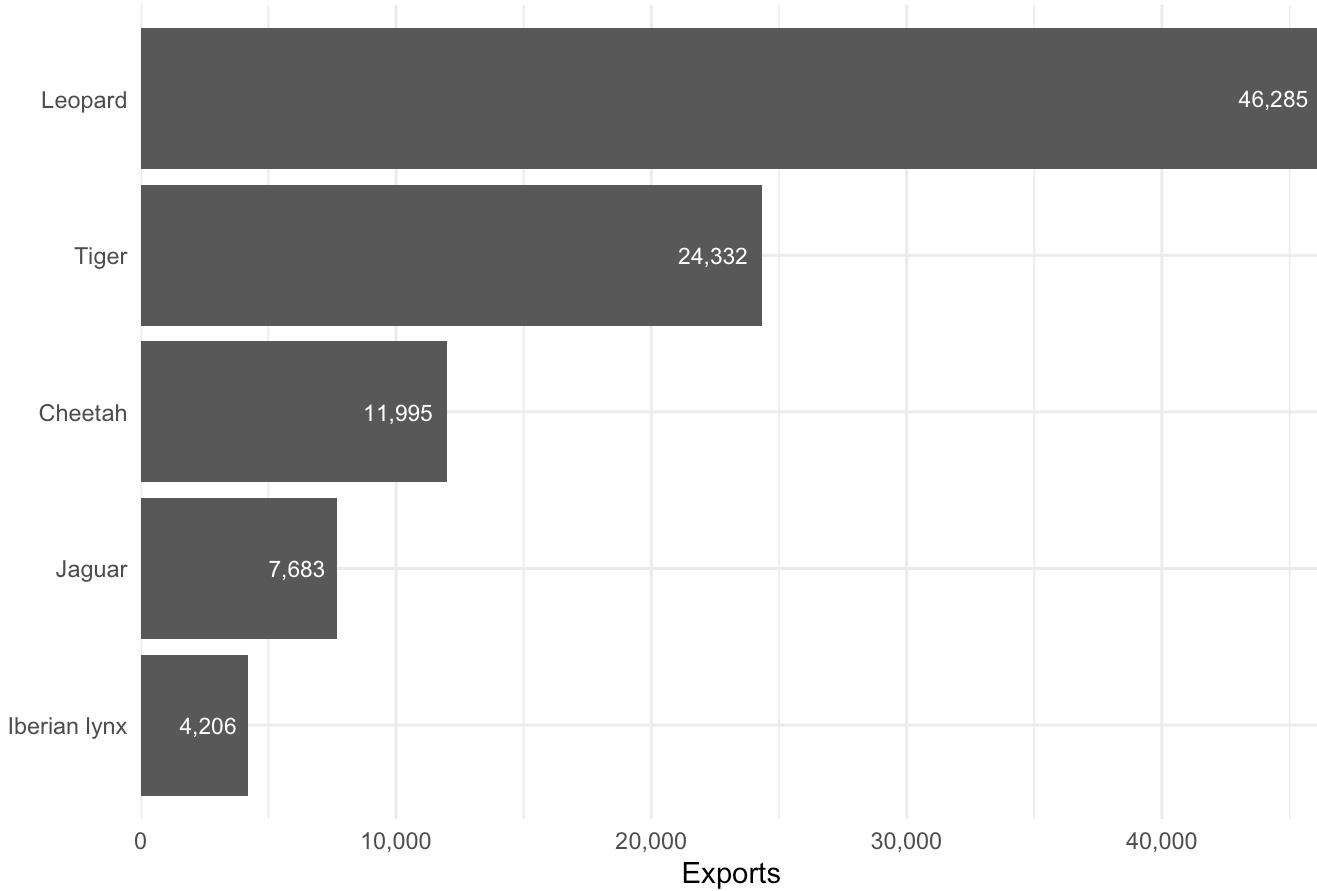
International Trade of Threatened & Endangered Big Cats



```
# Plot the top 5 most commonly exported cats
top_5 %>%
  ggplot(aes(x = reorder(common_name, n),
y = n)) +
  geom_col() +
  geom_text(aes(label = comma(n)),
hjust = 1.2,
color = "white", size = 3) +
  coord_flip() +
  labs(title = "International Trade of Threatened & Endangered Big Cats",
```

```
x = NULL,
y = "Exports") +
theme_minimal()+
scale_y_continuous(expand = expansion(mult = c(0,0)),
labels = comma)
```

International Trade of Threatened & Endangered Big Cats



Request for Feedback – I am not sure which of these I like better. I like the simplicity of the second plot but I also like how I can see the different types of exports that are being traded (specimen, live, trophy).

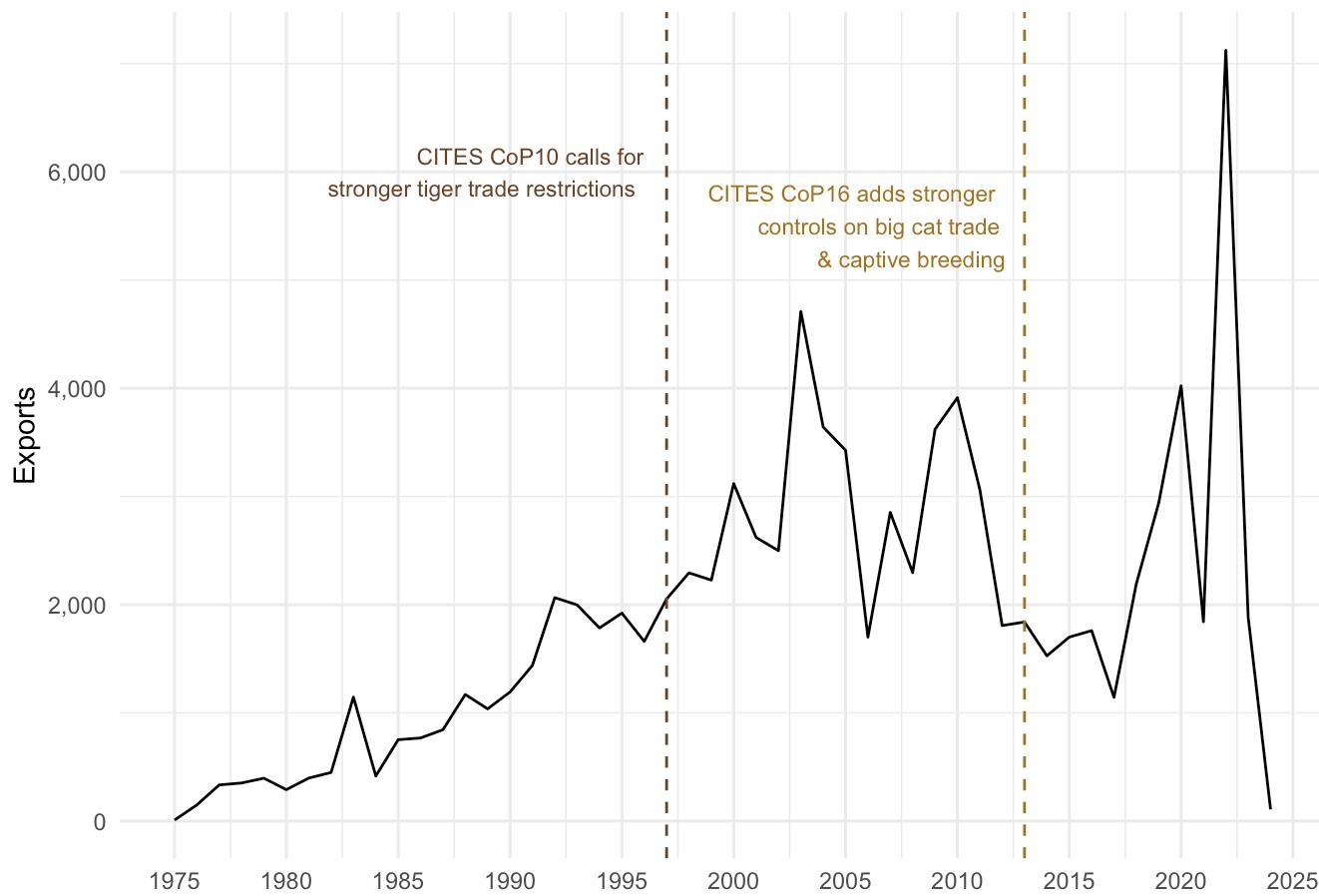
Across Time

```
# Aggregate data by year and taxon
gross_exports_aggregated <- gross_exports_clean %>%
  filter(taxon %in% c("Panthera tigris",
                      "Acinonyx jubatus",
                      "Panthera pardus",
                      "Panthera onca",
                      "Lynx pardinus")) %>%
  group_by(year, taxon) %>%
  summarise(total_count = sum(count), .groups = 'drop') %>%
  left_join(common_names, by = "taxon")%>%
  mutate(taxon = fct_reorder(taxon, total_count, .desc = TRUE))
```

```
# Total counts by year
counts_by_year <- gross_exports_aggregated %>%
  group_by(year) %>%
  summarise(total_count = sum(total_count))
```

```
ggplot(counts_by_year,
       aes(x = year,
            y = total_count)) +
  geom_line() +
  geom_vline(xintercept = 1997, linetype = "dashed", color = "#6B4226") +
  annotate(
    geom = "text",
    x = 1997, y = 6000,
    label = "CITES CoP10 calls for\nstronger tiger trade restrictions",
    size = 3, hjust = 1.1, color = "#6B4226") +
  geom_vline(xintercept = 2013, linetype = "dashed", color = "#A6761D") +
  annotate(
    geom = "text",
    x = 2013, y = 5500,
    label = "CITES CoP16 adds stronger\ncontrols on big cat trade\n& captive breeding",
    size = 3, hjust = 1.1, color = "#A6761D") +
  labs(
    title = "Exports of Big Cats Over Time",
    x = NULL,
    y = "Exports") +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma) +
  scale_x_continuous(breaks = seq(1975, 2025, by = 5))
```

Exports of Big Cats Over Time



Request for Feedback – I'm not sure the best way to visualize by data across time. I am struggling to know if I should aggregate across all the species or show it split up by species.

Across Exporters

```
export_country <- cites %>%
  filter(unit == "Number of specimens" | unit == "") %>%
  filter(taxon %in% c("Panthera tigris",
    "Acinonyx jubatus",
    "Panthera pardus",
    "Panthera onca",
    "Lynx pardinus")) %>%
  filter(term %in% c("skins", "trophies", "live", "specimens", "teeth")) %>%
  select(year, taxon,
    importer, exporter,
    importer_reported_quantity,
    exporter_reported_quantity, term)
```

```
# Summarize the trade quantity between each exporter-importer pair
links <- export_country %>%
  filter(year %in% 2020:2024) %>%
  group_by(importer, exporter) %>%
  summarize(value = sum(importer_reported_quantity, na.rm = TRUE)) %>%
```

```
ungroup() %>%
filter(value > 50) %>%
drop_na()
```

`summarise()` has grouped output by 'exporter'. You can override using the `groups` argument.

```
# Create a unique list of nodes (exporters and importers)
nodes <- data.frame(name = unique(c(links$exporter, links$importer)))

# Convert country names to indices for networkD3
links <- links %>%
  mutate(
    IDsource = match(exporter, nodes$name) - 1,
    IDtarget = match(importer, nodes$name) - 1
  )

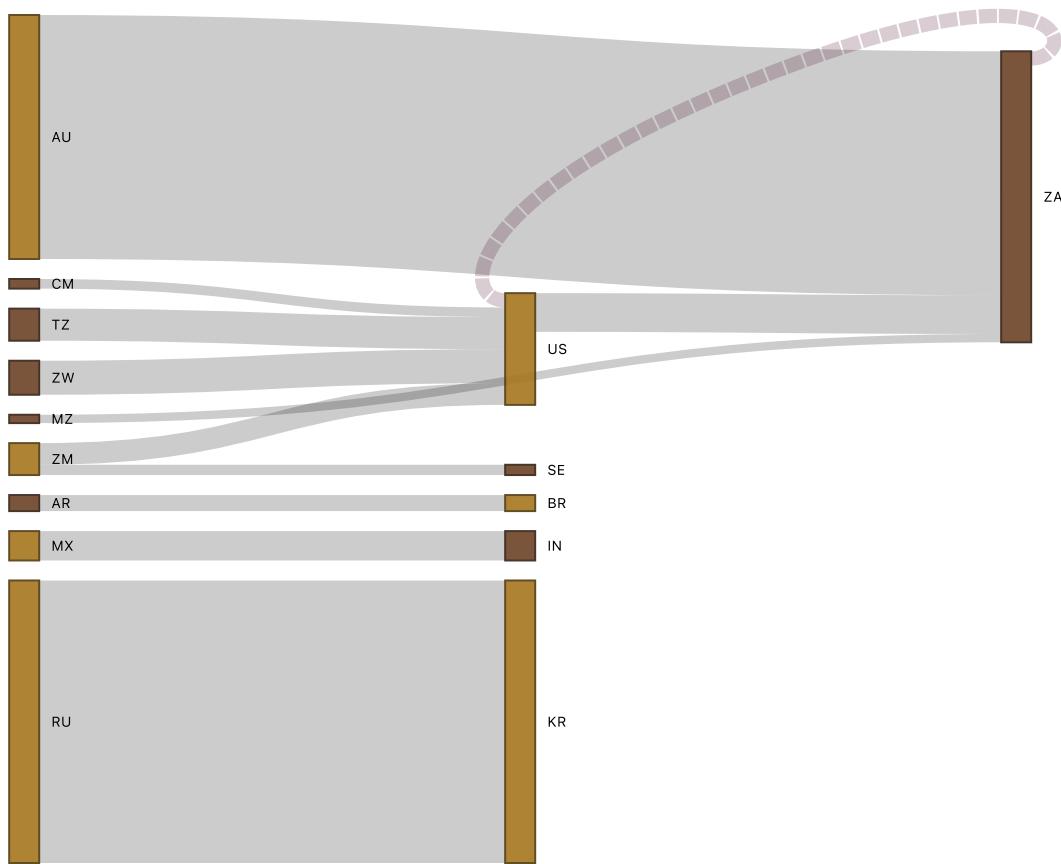
# Create the color scale for nodes
# You can modify this color scale to suit your needs (exporter = blue, importer = red)
my_color <- 'd3.scaleOrdinal() .domain(["exporter", "importer"]) .range(["#6B4226", "#A6734C"])

# Add a node type column to differentiate exporters and importers
nodes$type <- ifelse(nodes$name %in% links$exporter, "exporter", "importer")

# Apply color scale using the 'type' column
# Note that this uses the 'type' column to assign color
p <- sankeyNetwork(Links = links, Nodes = nodes,
                    Source = "IDsource", Target = "IDtarget",
                    Value = "value", NodeID = "name",
                    sinksRight = FALSE,
                    colourScale = my_color)
```

Links is a `tbl_df`. Converting to a plain data frame.

```
# Display the Sankey diagram
p
```



Request for feedback – how can I best incorporate color into my Sankey diagram? Right now I tried to do it by exporter and importer but I don't think I did it correctly.

Question 7

Answer the following questions:

What challenges did you encounter or anticipate encountering as you continue to build / iterate on your visualizations in R? If you struggled with mocking up any of your three visualizations (from #6, above), describe those challenges here.

- After reading a bit more into the metadata of on the CITES website I realized that I was misinterpreting the data and poorly analyzing it. Due to the nature of the wild animal trade, many different parts of an animal can be reported (i.e. fur, teeth, skin, or whole individual bodies such as trophies or live animals). For this reason, I realized I was incorrectly counting each reported entry as one individual which is not accurate. I spent a lot of time cleaning up my data set to get it into a format that accurately accounts for the correct number of reported specimens by species by year.
- I also struggled with making the sankey diagram, I wasn't sure the best way to filter down the data. I ended up displaying for the years 2020 to 2024 and for quantities over 50 individual animals. I ended up not making a map and rather showing the flow of species from exporting to importing countries.

What ggplot extension tools / packages do you need to use to build your visualizations? Are there any that we haven't covered in class that you'll be learning how to use for your visualizations?

- No?

What feedback do you need from the instructional team and / or your peers to ensure that your intended message is clear?

I would love more feedback on my Sankey diagram!

Works Cited

CITES Trade Database 2025. Compiled by UNEP-WCMC for the CITES Secretariat. Available at: trade.cites.org. Accessed February 19, 2025.