

Mini Project III - Tial and Jessica

```
#load libraries for extracting inat data
library(tidyverse)

-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   4.0.0     v tibble    3.3.0
v lubridate 1.9.4     v tidyr    1.3.1
v purrr    1.1.0

-- Conflicts -----
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become non-conflicting

library(rinat)

#extract raw data of all observations
raw_obs <- get_inat_obs_project("st-olaf-natural-lands", type = "observations")

9484 records

Getting records 0-200

Getting records up to 400

Getting records up to 600

Getting records up to 800
```

Getting records up to 1000

Getting records up to 1200

Getting records up to 1400

Getting records up to 1600

Getting records up to 1800

Getting records up to 2000

Getting records up to 2200

Getting records up to 2400

Getting records up to 2600

Getting records up to 2800

Getting records up to 3000

Getting records up to 3200

Getting records up to 3400

Getting records up to 3600

Getting records up to 3800

Getting records up to 4000

Getting records up to 4200

Getting records up to 4400

Getting records up to 4600

Getting records up to 4800

Getting records up to 5000

Getting records up to 5200

Getting records up to 5400

Getting records up to 5600

Getting records up to 5800

Getting records up to 6000

Getting records up to 6200

Getting records up to 6400

Getting records up to 6600

Getting records up to 6800

Getting records up to 7000

Getting records up to 7200

Getting records up to 7400

Getting records up to 7600

Getting records up to 7800

Getting records up to 8000

Getting records up to 8200

Getting records up to 8400

```
Getting records up to 8600
```

```
Getting records up to 8800
```

```
Getting records up to 9000
```

```
Getting records up to 9200
```

```
Getting records up to 9400
```

```
Getting records up to 9600
```

```
Done.
```

```
#Created a new data frame called "obs_clean" from the previous data frame "raw_obs" which contains the raw data.
obs_clean <- raw_obs |>

#Using the select() function, we selected specific columns we want to include in our new data frame.
  select(observed_on, description, latitude, longitude, species_guess, user_id, created_at, updated_at) |>

#We used the mutate() function to modify specific columns.
#The ymd() function converts date columns to follow this format: year-month-day
#For the "created_at" and "updated_at" columns, we used the str_extract() function to isolate the date portion of the column.
#The parse_number() function extracts numeric values from "latitude" and "longitude" columns
#Finally, the replace_na() function for the "species_guess" column to replace any "NAs" with "unknown".
  mutate(observed_on = ymd(observed_on),
         latitude = parse_number(latitude),
         longitude = parse_number(longitude),
         created_at = ymd(str_extract(created_at, "\\\\d{4}-\\\\d{2}-\\\\d{2}")),
         updated_at = ymd(str_extract(updated_at, "\\\\d{4}-\\\\d{2}-\\\\d{2}")),
         species_guess = replace_na(species_guess, "unknown")) |>

#Here, we used the rename() function to assign new names to specific columns.
  rename (species = species_guess,
         location = place_guess,
         quality = quality_grade,
         URL = uri,
         id_agreements = num_identification_agreements,
         id_disagreements = num_identification_disagreements)
```

```
#write_csv(obs_clean, "TJ_min13")
```

Variable Description Table : The variable descriptions come from the iNaturalist website: [https://www.inaturalist.org/terminology#num_identification_agreements]

variable name	data type	description
observed_on	Date	The date of when the observation was made
description	Character	Comments/notes from the observer
latitude	Numeric	The latitude of the location where the observation was made
longitude	Numeric	The longitude of the location where the observation was made
species	Character	The name of the observed taxon
user_id	Integer	The identifier for the observer
created_at	Date	The date of when the observation was uploaded onto iNaturalist
updated_at	Date	The date of when the uploader updated their post
location	Character	The location of where the observation was made
id_agreements	Integer	The number of matching identifications made by either the observer or community members
id_disagreements	Integer	The number of identifications made by either the observer or community members that do not match
quality	Character	Observation grades based on accuracy, completeness, and relevance. The grades are separated into three categories: “needs ID,” “casual,” and “Research Grade.” Observations are marked as “needs ID” if the community has not agreed on an identification, but the observation is verifiable. Observations are marked as “casual” if the observation is not verifiable, either due to missing or unreliable information. Observations are marked as “Research Grade” if the community agrees on the species identification.
URL	Character	The link/URL to the observation
captive	Logical	Uses “True” or “False” to indicate whether the observed organism was intentionally placed there by humans (true) or if it is a wild organism (false)

We have also created a function that allows the user to choose a location and species which will output a cleaned dataset of all the observations of that species at that location. Example below searches for all observations of blue jays at St. Olaf College.

```

inat_observations <- function(location, species) {
  raw_obs <- get_inat_obs_project(location, type = "observations")

  #Created a new data frame called "obs_clean" from the previous data frame "raw_obs" which
  obs_clean <- raw_obs |>

  #Using the select() function, we selected specific columns we want to include in our new data frame
  select(observed_on, description, latitude, longitude, species_guess, user_id, created_at)

  #We used the mutate() function to modify specific columns.
  #The ymd() function converts date columns to follow this format: year-month-day
  #For the "created_at" and "updated_at" columns, we used the str_extract() function to isolate the date
  #The parse_number() function extracts numeric values from "latitude" and "longitude" columns
  #Finally, the replace_na() function for the "species_guess" column to replace any "NAs" with "unknown"

  mutate(observed_on = ymd(observed_on),
         latitude = parse_number(latitude),
         longitude = parse_number(longitude),
         created_at = ymd(str_extract(created_at, "\\\d{4}-\\\d{2}-\\\d{2}")),
         updated_at = ymd(str_extract(updated_at, "\\\d{4}-\\\d{2}-\\\d{2}")),
         species_guess = str_to_lower(replace_na(species_guess, "unknown")))|>

  #Here, we used the rename() function to assign new names to specific columns.
  rename(species = species_guess,
         location = place_guess,
         quality = quality_grade,
         URL = uri,
         id_agreements = num_identification_agreements,
         id_disagreements = num_identification_disagreements) |>
  filter(species == {{species}})
  obs_clean
}

inat_observations(location = "st-olaf-natural-lands", species = "blue jay")

```

9484 records

Getting records 0-200

Getting records up to 400

Getting records up to 600

Getting records up to 800

Getting records up to 1000

Getting records up to 1200

Getting records up to 1400

Getting records up to 1600

Getting records up to 1800

Getting records up to 2000

Getting records up to 2200

Getting records up to 2400

Getting records up to 2600

Getting records up to 2800

Getting records up to 3000

Getting records up to 3200

Getting records up to 3400

Getting records up to 3600

Getting records up to 3800

Getting records up to 4000

Getting records up to 4200

Getting records up to 4400

Getting records up to 4600

Getting records up to 4800

Getting records up to 5000

Getting records up to 5200

Getting records up to 5400

Getting records up to 5600

Getting records up to 5800

Getting records up to 6000

Getting records up to 6200

Getting records up to 6400

Getting records up to 6600

Getting records up to 6800

Getting records up to 7000

Getting records up to 7200

Getting records up to 7400

Getting records up to 7600

Getting records up to 7800

Getting records up to 8000

Getting records up to 8200

Getting records up to 8400

Getting records up to 8600

Getting records up to 8800

Getting records up to 9000

Getting records up to 9200

Getting records up to 9400

Getting records up to 9600

Done.

	observed_on	description
1	2025-10-22	
2	2024-09-16	blue jay and white-breasted nuthatch
3	2023-11-20	Bird call, unsure of what kind.\n
4	2023-05-09	<NA>
5	2023-03-21	<NA>
6	2023-03-23	<NA>
7	2022-05-11	<NA>
8	2022-05-22	<NA>
9	2021-05-13	<NA>
10	2021-04-29	<NA>
11	2021-05-01	<NA>
12	2020-09-10	<NA>
13	2020-09-21	<NA>
14	2020-09-23	<NA>
15	2017-06-16	Blue Jay\r\nSt Olaf Natural Lands\r\nNorthfield, Minnesota
	latitude longitude species user_id created_at updated_at	
1	44.45973 -93.18333 blue jay 5648252	2025-10-22 2025-10-22
2	44.46429 -93.19521 blue jay 8578944	2024-09-16 2024-10-08
3	44.46128 -93.19084 blue jay 7403700	2023-11-20 2024-08-09
4	44.46445 -93.18122 blue jay 2473006	2023-05-23 2024-04-15
5	44.46224 -93.18612 blue jay 6516520	2023-03-21 2023-03-25

6	44.46460	-93.19078	blue jay	2473006	2023-04-18	2024-04-15
7	44.45863	-93.18149	blue jay	2473006	2022-05-11	2024-04-15
8	44.45436	-93.18884	blue jay	2473006	2022-05-23	2024-04-15
9	44.46470	-93.18113	blue jay	2473006	2021-05-13	2024-04-15
10	44.46699	-93.18192	blue jay	2473006	2021-04-29	2024-04-15
11	44.46649	-93.19199	blue jay	2473006	2021-05-01	2024-04-15
12	44.45878	-93.18093	blue jay	2473006	2020-09-11	2024-04-15
13	44.46385	-93.18251	blue jay	2473006	2020-09-22	2024-04-15
14	44.46223	-93.18427	blue jay	2473006	2020-09-24	2024-04-15
15	44.46361	-93.17831	blue jay	65504	2017-06-16	2023-07-08
				location	id_agreements	id_disagreements quality
1	St. Olaf College, Northfield, MN, US				1	0 research
2		Northfield, MN, US			2	0 research
3	St. Olaf College, Northfield, MN, US				3	0 research
4		Northfield, MN 55057, USA			1	0 research
5		Rice County, US-MN, US			2	0 research
6		Northfield, MN 55057, USA			2	0 research
7		Northfield, MN 55057, USA			2	0 research
8		Northfield, MN 55057, USA			3	0 research
9		Northfield, MN 55057, USA			2	0 research
10		Northfield, MN 55057, USA			2	0 research
11		Rice County, MN, USA			1	0 research
12		Northfield, MN 55057, USA			1	0 research
13		Northfield, MN 55057, USA			1	0 research
14		Northfield, MN 55057, USA			1	0 research
15	Northfield, Minnesota, United States				3	0 research
				URL	captive	
1	https://www.inaturalist.org/observations/322475164				FALSE	
2	https://www.inaturalist.org/observations/242103193				FALSE	
3	https://www.inaturalist.org/observations/191608899				FALSE	
4	https://www.inaturalist.org/observations/163357730				FALSE	
5	https://www.inaturalist.org/observations/151872743				FALSE	
6	https://www.inaturalist.org/observations/155543120				FALSE	
7	https://www.inaturalist.org/observations/116613887				FALSE	
8	https://www.inaturalist.org/observations/118325658				FALSE	
9	https://www.inaturalist.org/observations/78590666				FALSE	
10	https://www.inaturalist.org/observations/75652549				FALSE	
11	https://www.inaturalist.org/observations/76139375				FALSE	
12	https://www.inaturalist.org/observations/59185992				FALSE	
13	https://www.inaturalist.org/observations/60447292				FALSE	
14	https://www.inaturalist.org/observations/60645855				FALSE	
15	https://www.inaturalist.org/observations/6685055				FALSE	