Testing for Tolerance in a Box Experiment in Wild Vervet Monkeys - M.AungKyaw - 2022-2023

juin 03, 2024 16:03

Table of Contents

Introduction 1
0.Opening the data 1
Loading data 1
1.Explore the data 1
Description of the initial datset - “Boxex” 1
Variables of Boxex 1
2. Treating missing data 1
2.1. Creating a new dataframe - Bex 1
2.1.1 Merging Male placement corn and MaleCorn 1
2.1.2 Cleaning FemaleCorn 1
2.2 Cleaning variables with missing data 1
2.3 Treating variables with missing data 1
2.3.1 Cleaning “Remarks” - (2181 NA’s) 1
2.3.2 Cleaning “Intruder ID” - (2737 NA’s) 1
2.3.3 Cleaning “IDIndividual1” - (2143 NA’s) 1
2.6 Cleaning “Audience” - (924 NA’S) 1
2.7 Cleaning “OtherResponse” - (2758 NA’S) 1
2.8 Cleaning of “Time” 1
2.9 Cleaning DyadDistance 1
2.10 Cleaning Female and Male ID 1
2.12 Dyad Response (7) 1
2.13 Final check: remaing NA’s in Bex? 1
3. Correction and creation of New Variables 1
3.1 Making a backup of Bex 1
3.2 Treating Remarks before processing with new modifications 1
3.2.2 Vizualization of the Remarks Keywords 1
3.2.3 Exporting of Bex 1

3.2.4 Journal of manual changes in Bex excel spreadsheet ... 1
UPDATES FROM HERE ... 1
3.3 Re Uploading the dataset after the treatment of the Remarks ... 1
Treatment and cleaning of last variables before stats!? ... 1
4.2 Time - Creation of Period and Hour ... 1
Date - Creation of Month and Day ... 1
Group - Ok ... 1
Male and Female ID - Creation of Dyad, Trial and Session ... 1
Female Corn and Male Corn ... 1
DyadDistance - Creation of proximity ... 1
Code to keep and place again ... 1
Code to settle ... 1
Exploratory Graph (To organise) ... 1
Unique values in FemaleID ... 1
Sections below are here for the organization of my paper and will be worked on once the data cleaning and exploration is done ... 1
3. Describing the data ... 1
4.Visualizing the data ... 1
5.Research question & Hypothesis ... 1
Research question ... 1
Hypothesis ... 1
6.Statistical tests and analisis of the data ... 1
Statistical tests ... 1
7. Plotting the results of the analysis ... 1
9. Interpretation of the results ... 1
10. Comeback on the research question and hypothesis ... 1
11. Bibliography ... 1
12. Organization for my paper ... 1
Glossary ... 1
Bibliography ... 1
Annex ... 1
Annex 1 : View of the dataset when imported - First 6 entries of each variable ... 1

Introduction

0.Opening the data

Loading data

- First I downloaded the **knitr package** to create outputs as html, pdf or word files when knitting my r markdown file. I also loaded the **pander** package for better presentation
- The **dplyr** package was installed for better manipulation of the data as filtering or creating new variables and **lubridate** for a better manipulation of dates and time
- Then, I installed the **readxl package** to import my dataset which is called **Box Experiments.xls**
- This dataset contains information related to my master thesis project. I used cyber tracker in order to record the behaviors of dyads of Vervet monkeys in a box experiment on tolerance from September 2022 to September 2023

1.Explore the data

Description of the initial datset - "Boxex"

```
## Glimpse of the Box Experiment dataset:

## Rows: 2,795
## Columns: 20
## $ Date                  <dttm> 2022-09-27, 2022-09-27, 2022-09-27, 2022-
09-27,…
## $ Time                  <dttm> 1899-12-31 09:47:50, 1899-12-31 09:50:07,
1899-…
## $ Data                  <chr> "Box Experiment", "Box Experiment", "Box
Experim…
## $ Group                 <chr> "Baie Dankie", "Baie Dankie", "Baie Dankie",
"Ba…
## $ GPSS                  <chr> "-28.010549999999999", "-
28.010549999999999", "-…
## $ GPSE                  <chr> "31.191050000000001", "31.191050000000001",
"31.…
## $ MaleID                <chr> "Nge", "Nge", "Nge", "Nge", "Nge", "Nge",
"Nge",…
## $ FemaleID              <chr> "Oerw", "Oerw", "Oerw", "Oerw", "Oerw",
"Oerw", …
## $ `Male placement corn` <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, …
## $ MaleCorn              <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
3, …
## $ FemaleCorn            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, …
## $ DyadDistance          <chr> "2m", "2m", "1m", "1m", "0m", "0m", "0m",
```

```
"0m", …
## $ DyadResponse          <chr> "Tolerance", "Tolerance", "Tolerance",
"Toleranc…
## $ OtherResponse         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, …
## $ Audience              <chr> "Obse; Oup; Sirk", "Obse; Oup; Sirk", "Oup;
Sirk…
## $ IDIndividual1         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, …
## $ IntruderID            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
"Sey…
## $ Remarks               <chr> NA, NA, "Nge box did not open because of the
bat…
## $ Observers             <chr> "Josefien; Michael; Ona; Zonke", "Josefien;
Mich…
## $ DeviceId              <chr> "{7A4E6639-7387-7648-88EC-7FD27A0F258A}",
"{7A4E…
```

- I am now using the **View** function to have a sight on the entire dataset and **glimpse** to display a summary of my dataset
- I have **20 variables** (here columns) and **2795 trials** (here rows)
- I will now make a brief summary of each variables and their use before creating a new dataframe (df) with my variables of interest that I will call **Bex**
- The highlighted variables are the ones I will use for **Bex**. I will then **clean the data** before heading to the **statistical analysis** and the **interpretation of the results**

Variables of Boxex

- **Date** : "Date" is in a **POSIXct** format which is appropriate for the display of time
  - I want to use the date to know **how many sessions** have been done with each dyads in my experiment.
  - I will create a variable called **Session** where **1 session = 1 day**
  - The data has values from the **14th of September 2022** until the **13th of September 2023**
  - I may consider separating the **12 months** of data in **4 seasons** to make a preliminary check of a potential effect of seasonality. Nevertheless since we did not use any tools to measure the weather, temperature, humidity or food availability (also related to seasonality and weather). Categorizing my data in 4 without having further data would then be quite arbitrary. If I end up doing it in my report, it will be done without any intention to include it in my scientific analysis nor my scientific report .
- **Time** : "Time is coded" in a **POSIXct** format
  - I do not plan to use this variable but we can see that "Time" has the correct hours displayed with a date which is incorrect.

  - (In the case I wanted to observe **when the trials occurred during the day** as time may have an influence on their behavior (**Isbell & Young 1993**)) I would need to correct the incorrect display of the date in the dataset.
  - This variable could also be useful to see when the **seasonal effect** took place as we only went in the morning during summer because of the heat while we went later and for longer times in the field to do the box experiment in winter
  - For now, the values in "Time" are all on the same (wrong) day which is the **31st of December**
  - Note: I first did not intend to keep **Time** in Bex but I needed this variable to see the order of the trails within a day. I finally decided to keep it.
- Data : chr "Data" is coded as **character**
  - It describes **the type of data** being recorded in the software **cybertracker**. We installed the software on tablets to record the different behaviors of vervet monkeys in our research center
  - In our case, my data was recorded in cybertracker as **Box Experiment** as we created a form specifically for this experiment
  - For this reason we can remove this column since the information it contains is unecessary and redundant
- Group : chr The data is coded in r as a **character**
  - It describes the **group of monkey** in which we did the trial
  - I will keep this column to see the amount of trials that we did in the 3 group of monkeys which are Baie-Dankie **(BD)**, Ankhase **(AK)**, and Noha **(NH)**
- GPSS : num "GPSS" is coded as **numerical**
  - It gives the **south coordinates** in which we started the experiment
  - I do not plan to use coordinates nor look at locations so I will remove this acolumn
- GPSE : num "GPSE" is coded in as **numerical**
  - It gives the **east coordinates** in which we started the experiment
  - I do not plan to use coordinates nor look at locations so I will remove this column
- **MaleID** : chr "MaleID" is coded as **character**
  - It indicates the **name of the male involved in the trial**
  - I plan to use this to see how factors related to the individual may influence the experiment (age, sex, rank)
  - It will also help me see which behaviour was displayed by each individuals (here males)
- **FemaleID** : chr "FemaleID" is coded as **character**
  - It indicates the **name of the female involved in the trial**
  - I plan to use this variable in the same way as "Male ID"

  - It will also help me see which behaviour was displayed by each individuals (here females)
- **Male placement corn**: dbl "Male placement corn is coded in r as **double**
  - It gives the **amount of corn given to the male of the dyad before the trials**
  - Within a session it happened that we gave more placement corn to attract the monkeys again to the boxes. This lead to an update of the number in the same session. The number found at the end of the session is the total placement corn an individual has received
  - I will fuse this column with **male corn** as the data has been separated between these two variables. This is due to a mistake when creating the original box experiment form in cybertracker
  - This variable could be related to the level of motivation of a monkey but as it is not directly related to my hypothesis I may not use this column. I will re-consider the use of this column later on
  - In regards of this possibility I will change the format of the variable to numerical
- **MaleCorn** : dbl "MaleCorn" is coded in r as **double**
  - It gives the same information as in **male placement corn**
  - I will import the values from "male placement corn" into this one
  - I will change the format of the variable to numerical
- **FemaleCorn** : dbl The data is coded in r as **double**
  - It gives the **amount of corn given to the female of the dyad before the trials**
  - It works in the same way as "male placement corn"/"MaleCORN"
  - I will change the format of the variable to numerical
- **DyadDistance** : chr The data is coded in r as **character**
  - It gives the **distance for each trial** that we have done with the dyads.
  - The trial number 1 for each dyad was at 5 meters.
  - The maximum was around 10 m while the minimum is 0
  - We will have to remove the "m" for meters in order to have a numerical variable instead of character
  - Also, since the very first trials per dyad can be considered as a kind of learning phase, i may remove the **15 first trials** that were made for each dyad
- **DyadResponse** : chr The data is coded in r as **character**
  - It indicated which **behaviour was produced by the dyad's during each trial**

- The different behaviours were: **Distracted**, **Female aggress male**, **Male aggress female**, **Intrusion**, **Loosing interest**, **Not approaching**, **Tolerance** and **Other**
- I will change the columns associated to each behavior (i.e. Response) of **DyadResponse** into dichotomic variables in order to see the frequency of each behaviour
- This will allow me to see which behavior occurred more ,and behavioural differences could be found between dyads
- As multiple response could occur within the same trial, multiple behaviors can be found in a single cell. I will create a hierarchy to reduce the amount of behaviors assigned to each trial (if there is more than one). This will also be complemented with the information found in the column **remarks**
    1. correct any mistakes (ex. if tolerance and aggression are together aggression>tolerance)
    2. assign as few labels per trial
    3. get a better View and understanding of the data and the most common behaviours produced by each dyad
    4. create variables that can complement the behaviour found (ex. not approaching + looks at partner would be looks at partner + a new variable called hesistant to see when the did not come but look at the other individual /)
- Projection of the hierarchy (changes will be made)
    - Create a table with each combination existing
    - Decide what is more important
    - Ex:
        - Aggression > Tolerance
        - Tolerance > Not approaching -> Create a variable called hesistant in addtion to the tolerance count to see frequency of tolerance behaviour that happened after > 1min
        - Tolerance > Loosing interest
        - Tolerance > Intrusion
        - Not approaching = looking box but not coming while Loosing interest = not paying attention to the box
        - Intrusion > Loosing interest
        - Intrusion > Not approaching
        - Not approaching > Looks at partner
        - We can code every look at partner as no approaching and keep the count of looks at partner as additional information
        - Not approaching >?> Loosing interest ? !!
        - Define distracted
        - Not approaching > Distracted

    - Aggression > Not approaching
    - Other > Look case by case and categorize depending of behavior
    - Remarks may be used for the same reason
- **OtherResponse** : chr "The data"OtherResponse" is coded as **character**
    - It describes **any behaviour that is different from the ones found in Dyad Response** (meaning ≠ tolerance, aggression, intrusion, loosing interest, not approaching, distracted, looks at partner that where categorized as **other**)
    - I will have to look at every **OtherResponse** and rename each entry in one of the response already if existing. I will proceed case by case.
    - If I want to do an intermediate manipulation I may rename every NA in "OtherResponse" into **Response** to see the amount of case to treat and how many occurrences seem to not fit in the categories of "DyadResponse"
- **Audience** : chr "Audience" is in r as **character**
    - It gives the **names of the individuals in the audience**
    - I would like to use it to see the **amount of audience (big vs small)** and the **dominance level of the audience (high vs low)**
    - I will create a variable called **NAudience** to see hoy many individuals are in the audience for each trial
    - After calculating the elo ratings of the individuals using another dataset (Life history), I will create a dichotomic variable called **RankAudience** to see effects related to rank with the effect of audience
- **IDIndividual1** : chr "IDIndividual1" is coded in r as **character**
    - It gives the **names of the individuals that did not approach, showed aggression, got distracted or lost interest** during a trial
    - I will have to look at it to see how often these behaviors occurred
    - I will consider how to use this variable during the cleaning of the data
- **IntruderID** : chr "IndtruderID" is coded as **character**
    - It gives the **name of the individual that intruded the experiment during a trial**
    - Intrusion could mean, invade the space of the experiment and interact with one of our individual, steal the food, show agnostic behavior, stand in very close proximity of the dyad's individuals
- **Remarks** : chr The data is coded in r as **character**
    - It gives either additional information concerning the experiment when unusual behaviors occurred , mistakes that needed to be corrected or details that we wanted to record in case we would need them
- Observers :chr The data is coded in r as **character**
    - It gives the **names of the observers during the experiment**

  - We will not use this data as we do not look at the effect that an experimenter would have on the monkeys
  - (Should I still look at an effect of the amount of experimenter?...maybe better for detailled analysis of our study)
- DeviceID :chr "The data"DeviceID" is coded in r as **character**
  - It gives the **name of the device/tablet** used to record the data during the experiment
  - We will not use this data either

2. Treating missing data

2.1. Creating a new dataframe - Bex

- Since I do not want to work with the whole dataset, I'm gonna select the variables of interest using the function **select**
- I will keep Time, Date, Group, MaleID, FemaleID, MaleCorn, Male placement corn, FemaleCorn, DyadDistance, DyadResponse, OtherResponse, Audience, IDIndividual1, IntruderID, Remarks

```
## Rows: 2,795
## Columns: 15
## $ Time                  <dttm> 1899-12-31 09:47:50, 1899-12-31 09:50:07,
1899-…
## $ Date                  <dttm> 2022-09-27, 2022-09-27, 2022-09-27, 2022-
09-27,…
## $ Group                 <chr> "Baie Dankie", "Baie Dankie", "Baie Dankie",
"Ba…
## $ MaleID                <chr> "Nge", "Nge", "Nge", "Nge", "Nge", "Nge",
"Nge",…
## $ FemaleID              <chr> "Oerw", "Oerw", "Oerw", "Oerw", "Oerw",
"Oerw", …
## $ MaleCorn              <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
3, …
## $ `Male placement corn` <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, …
## $ FemaleCorn            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, …
## $ DyadDistance          <chr> "2m", "2m", "1m", "1m", "0m", "0m", "0m",
"0m", …
## $ DyadResponse          <chr> "Tolerance", "Tolerance", "Tolerance",
"Toleranc…
## $ OtherResponse         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, …
## $ Audience              <chr> "Obse; Oup; Sirk", "Obse; Oup; Sirk", "Oup;
Sirk…
## $ IDIndividual1         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, …
```

```
## $ IntruderID            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
"Sey…
## $ Remarks               <chr> NA, NA, "Nge box did not open because of the
bat…
```

2.1.1 Merging Male placement corn and MaleCorn

- I want to process all the missing data in Bex. But before, I will merge the column **MaleCorn** and **Male placement corn** as the data of both columns is supposed to be together under "MaleCorn"
- Looking manually in the Bex table it seems that very few data is in **MaleCorn** while most of it seems to be in **Male placement corn**
- Every time there is a missing value in Male placement corn we can see a value in Male Corn, I will then create a new variable MaleCorn where every time that there is NA in male placement corn the value will be taken in MaleCornOld (previous malecorn). If there is no NA it will take the value of ´Male placement corn´
- I will first **rename MaleCorn to MaleCornOld**, then **check the amount of NA's** and then **merge "MaleCornOld" and "male placement corn"** into the **new variable "MaleCorn"**

```
## Number of rows with common NAs in MaleCornOld and 'Male placement corn':
1499
## Number of occurrences of 0 in MaleCorn: 1499
## Number of remaining NA values in MaleCorn: 0
```

- I have found **1499 NA in common** between MaleCornOld and 'male placement corn', **1609 NA in Male placement corn** and **2685 in MaleCorn old**

- For the **merge of MaleCornOld and Male placement corn**, I used different conditions: 1.In this code, a new variable MaleCorn is created. If there is a missing value in Male placement corn, it takes the corresponding value from MaleCornOld; otherwise, it takes the value from Male placementcorn. 2.If there are no value in both MaleCornOld and Male placement corn (NA,NA) for a given row, I would like the code to display 0 as it means that no placement was given

- In this way, I should not loose any data, minimize the mistakes and already transform the NA's of this variable into a number which will remove the remaining NA's which are meant to be 0

- After the merge I found that there were **no NA's remaining** in the **"New" Male Corn** and that **1499 0's** where found in the column which **corresponds to the amount of common NA's found previously** between the **"Old" Male Corn** and **male placement corn**

2.1.2 Cleaning FemaleCorn

```
## Number of remaining NA values in FemaleCorn: 0
```

2.2 Cleaning variables with missing data

- Now in order to see where are located the missing points in the data, I'm going to **print** the variables **with and without NA's**

- The function **sapply** is used to apply the function **sum** for NA's to each column of the data frame, so each variable

```
## Variables with Missing Data:
```

	x
MaleID	19
FemaleID	60
DyadDistance	33
DyadResponse	47
OtherResponse	2758
Audience	924
IDIndividual1	2143
IntruderID	2737
Remarks	2181

```
## Variables with No Missing Data:
```

	x
Time	0
Date	0
Group	0
FemaleCorn	0
MaleCorn	0

- We can see that out of the 14 variables we have in **Bex** we have **9 variables with missing data** which are **Male ID, Female ID, DyadDistance, DyadResponse, OtherResponse, Audience, IDIndividual1, IntruderID, Remarks**: I will proceed to clean these variables one by one
- MaleID 19
- FemaleID 60
- DyadDistance 33
- DyadResponse 47
- OtherResponse 2758
- Audience 924
- ID Individual1 2143
- IntruderID 2737
- Remarks 2181

- Before making treating the NA's in the dataset I will make a backup of the data at this point:

2.3 Treating variables with missing data

2.3.1 Cleaning "Remarks" - (2181 NA's)

- Since most of the time we did not have any remarks it is understandable that this variable contains 2181 NA's out of 2795 rows
- I will first transform every missing data in the column Remark into **No Remarks** and then check that the amount of "No remarks" found
- After the changes we can effectively see that we have **2181 "No Remarks"** and we have no missing data left in that column, I will treat this column by hand once all the NA's have been removed from the dataset

```
## Number of 'No Remarks' in the 'Remarks' column: 2181

##
## No Remarks    Remarks
##       2181        614
```

2.3.2 Cleaning "Intruder ID" - (2737 NA's)

- **Intruder ID** is a variable that contains the **name of the individuals that made and intrusion during a trial**.
- If more than one individual intruded, his name may be in the comments, which I will check when treating the data from this column
- Because nothing was entered when there was no intrusion, I will replace every NA's by **No Intrusion**
- Also, I will use a function to create a new dichotomic variable called **Intrusion**. Every time there is a value in IntruderID, it should display 1 (Yes), if not a 0 (No intrusion)

```
## Number of 'No Intrusion' in the 'Intruder ID' column after replacement:
2737
```

- We previously had 2737 NA's in IntruderID while now we have the same amount of occurrences of IntruderID which shis that the transformation went as intended

2.3.3 Cleaning "IDIndividual1" - (2143 NA's)

- IDIndividual1 is meant to report the name of the individual that did a behavior such as not approach, show aggression or loose interest during a trial
- I will now replace every NA in this column by **No individual** and print the amount of NA's left and the amount of changes made

```
## Number of NAs replaced in IDIndividual1: 2143
## Number of remaining NA values in IDIndividual1: 0
```

2.6 Cleaning "Audience" - (924 NA'S)

- Audience is made to report every name of individuals around our dyad during a given trial
- I will replace every NA by **No audience** as no entry means the absence of other individuals around
- I will also create a new variable called "Amount audience" that will have to tell me how many individuals are found in the column Audience

```
## Number of changes made in 'Audience': 924

## Remaining NA values in 'Audience': 0
```

2.7 Cleaning "OtherResponse" - (2758 NA'S)

```
## Number of changes made in 'OtherResponse': 2758

## Remaining NA values in 'OtherResponse': 0
```

2.8 Cleaning of "Time"

- Since the reading of the data is more complicated without the time, which was usefull to know which trial was before or after, I changed the code made for Bex and added **Time** in the dataframe. Since I will need it for the cleaning of Dyaddistance, I will now extract the time from the date. Even if the date is wrong as seen in the first output, the time is correct. As in the second output, only the time has been kept

```
## [1] "1899-12-31 09:47:50 UTC" "1899-12-31 09:50:07 UTC"
## [3] "1899-12-31 09:53:11 UTC" "1899-12-31 09:54:28 UTC"
## [5] "1899-12-31 09:55:19 UTC" "1899-12-31 09:56:56 UTC"

## [1] "09:47:50" "09:50:07" "09:53:11" "09:54:28" "09:55:19" "09:56:56"
```

2.9 Cleaning DyadDistance

- Before looking at the NA's of Dyaddistance I will remove the "m" that is in front of every number to have a numerical variable
- Then I will look at the location of the NA's in the data to treat them case by case.

```
## Warning: NAs introduits lors de la conversion automatique

## # A tibble: 69 × 16
##    Time     Date                Group    MaleID FemaleID FemaleCorn
DyadDistance
##    <chr>    <dttm>              <chr>    <chr>  <chr>         <dbl>
<dbl>
##  1 12:09:34 2022-09-27 00:00:00 Baie Da… Xia    Piep              7
NA
##  2 12:13:28 2022-09-27 00:00:00 Baie Da… Xia    Piep              7
NA
##  3 16:02:32 2022-09-15 00:00:00 Ankhase  Sho    Ginq              6
NA
##  4 10:46:33 2023-08-17 00:00:00 Baie Da… Xia    Piep              0
NA
##  5 09:30:17 2023-07-29 00:00:00 Baie Da… Xin    Ouli              0
```

```
NA
##  6 12:08:51 2023-07-11 00:00:00 Baie Da… Xia    Piep              0
NA
##  7 13:30:07 2023-06-29 00:00:00 Baie Da… Sey    Sirk              0
NA
##  8 09:54:24 2023-06-27 00:00:00 Ankhase  Sho    Ginq              0
NA
##  9 10:13:56 2023-06-23 00:00:00 Ankhase  Sho    Ginq              0
NA
## 10 09:39:04 2023-06-15 00:00:00 Ankhase  Sho    Ginq              2
NA
## # ℹ 59 more rows
## # ℹ 9 more variables: DyadResponse <chr>, OtherResponse <chr>, Audience
<chr>,
## #   IDIndividual1 <chr>, IntruderID <chr>, Remarks <chr>, MaleCorn <dbl>,
## #   Intrusion <dbl>, AmountAudience <dbl>
## Number of NA values in DyadDistance column (using second approach): 69
## Rows with NA values in DyadDistance column: 24, 27, 95, 492, 744, 971,
1113, 1130, 1164, 1261, 1341, 1396, 1491, 1583, 1683, 1693, 1717, 1718, 1719,
1724, 1725, 1739, 1755, 1756, 1757, 1764, 1779, 1782, 1792, 1799, 1800, 1840,
1841, 1868, 1869, 1888, 1891, 1892, 1896, 1911, 1912, 1915, 1918, 1919, 1952,
1953, 1958, 1980, 1981, 1984, 1986, 1996, 2000, 2009, 2054, 2104, 2105, 2191,
2233, 2234, 2287, 2437, 2569, 2579, 2580, 2643, 2676, 2709, 2729
```

- We have 69 missing values in DyadDistance. I will look at each row in it's context as the actual distance of the box was always dependent of the previous trials. I will start with the bigger number as for now the oldest trial is at the last row while the closest one is in row 1.
    - If **tolerance** was achieved **twice in a row** = <1m
    - If **aggression** (male agress female or female agress male), not approaching,or **loosing interest** occured = >1m
    - If **distracted** or **intrusion** occured = same distance

    1. **24** - In trial23 (0m) there was aggression then at trial24 (1m) there was tolerance. The 24th trial is supposed to be at **1m**
    2. **27** - In trial25 (1m) there was not approaching then at trial26 (2m) there was tolerance.The 25th trial is supposed to be at **2m**
    3. **95** - In trial93 (2m) there was male agress female then at trial 94 (3m) there was not approaching. The 95th trial is supposed to be at **4m**
    4. **492** - In trial490 (0m) there was tolerance then at trial491 (0m) there was tolerance. The 492nd trial is supposed to be at **0m**
    5. **744** - In trial742 (3m) there was aggression then at trial743 (4m) there was tolerance. The 744th trial is supposed to be at **4m**
    6. **971** - In trial969 (0m) there was tolerance then at trial970 (0m) there was tolerance. The 971st trial is supposed to be at **0m**

7. **1113** - In trial1111 (2m) there was tolerance then at trial1112 (0m) there was tolerance. The 1113th trial is supposed to be at **0m**
8. **1130** - In trial1128 there was another dyad so we can not use this cell. Then at trial1129 (3m) there was not approaching. Nevertheless, we don't have any DyadResponse, i will thus **delete this row**
9. **1164** - In trial1162 (3m) there was not approaching then at trial1163 (3m) there was not approaching. The 1164th trial is supposed to be at **4m**
10. **1261** - The two preivous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
11. **1341** - In trial1339 (0m) there was tolerance then at trial1340 (0m) there was tolerance. The 1341st trial is supposed to be at **0m**
12. **1396** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
13. **1491** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
14. **1583** - In trial1581 (2m) there was not approaching and intrusion then at trial1582 (2m) there was not approaching. The 1583rd trial is supposed to be at **3m**
15. **1683** - One trial only was made with tolerance (2m) but since there are no DyadResponse I will **delete this row**
16. **1693** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
17. **1717** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
18. **1718** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
19. **1719** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
20. **1724** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
21. **1725** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
22. **1739** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
23. **1755** - since there are no DyadResponse I will **delete this row**
24. **1756** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
25. **1757** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
26. **1764** - Since there are no DyadResponse I will **delete this row**
27. **1779** - It seems like it was the first trial of the Dyad Pom Xian, if so, the distance has to be **5m**

28. **1782** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
29. **1792** - Trial1791 was intrusion (4m) so this trial should be at **4m**
30. **1799** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
31. **1800** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
32. **1840** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
33. **1841** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
34. **1868** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
35. **1869** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
36. **1888** - he two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
37. **1891** - Since there are no DyadResponse I will **delete this row**
38. **1892** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
39. **1896** - Since there are no DyadResponse I will **delete this row**
40. **1911** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
41. **1912** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
42. **1915** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
43. **1918** - In trial1916 (4m) there was tolerance then at trial1917 (4m) there was not loosing interest The 1918th trial is supposed to be at **4m**
44. **1919** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
45. **1952** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
46. **1953** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
47. **1958** - In trial1956 (2m) there was tolerance then at trial1957 (2m) there was distracted. The 1958th trial is supposed to be at **2m**
48. **1980** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
49. **1981** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
50. **1984** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**

51. **1986** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
52. **1996** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
53. **2000** - In trial1997 and 1999 (5m) there was tolerance then at trial1999 (5m) there was intrusion. The 2000th trial is supposed to be at **4m**
54. **2009** - Since there are no DyadResponse I will **delete this row**
55. **2054** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
56. **2104** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
57. **2105** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
58. **2191** - In trial2189 (1m) there was not approaching then at trial2190 (2m) there was not approaching. The 2191st trial is supposed to be at **3m**
59. **2233** - In trial2231 (3m) there was not approaching then at trial2232 (4m) there was not approaching. The 2233rd trial is supposed to be at **5m**
60. **2234** - The trial did not happen because they where not at the right distance. I will thus **delete this row**
61. **2287** - Since there are no DyadResponse I will **delete this row**
62. **2437** - Since there are no DyadResponse I will **delete this row**
63. **2569** - In trial2567 (1m) there was tolerance then at trial2568 (1m) there was tolerance. The 2569th trial is supposed to be at (0m)
64. **2579** - In trial2577 (1m) there was tolerance then at trial2578 (0m) there was not approaching. The 2579th trial is supposed to be at **1m**
65. **2580** - The two previous trials were made with another Dyad. Also DyadResponse is not available. I will thus **delete this row**
66. **2643** - Since there are no DyadResponse I will **delete this row**
67. **2676** - In trial2674 (1m) there was tolerance then at trial2675 (0m) there was tolerance. The 2676th trial is supposed to be at **0m**
68. **2709** - In trial2707 (2m) there was tolerance then at trial2708 (2m) there was tolerance. The 2709th trial is supposed to be at **1m**
69. **2729** - In trial2727 (3m) there wastolerance then at trial2728 (2m) there was tolerance. The 2729th trial is supposed to be at **2m**

- Now that I have looked at each missing line and saw which ones to keep, I decided to create a new variable called **Distance**. I will also to create a new variable called **No trial**.

- For the variable **Distance** I will replace each row where there was missing data with a value and I will delete the ones where no values could be assigned. This will allow me to have no missing data and find a number to each trial that has been done

- Before making the changes i'm gonna make a backup called **BackupbeforeDistanceNA**

```
## Number of NA's in DyadDistance after replacements and deletions: 1
## Data size after deletions: 2748

## Row index with NA in DyadDistance: 1925
```

*It seems that there is still the row 1925 with an NA in DyadDistance

70. **1925** - In trial1923 (2m) there was distracted then at trial1924 (2m) there was tolerance. The 2725th trial is supposed to be at **2m**

```
## Row index with NA in DyadDistance:
```

*In this modification, I added a check to see if the columns Dyadistance and Distance already exist in your dataframe (Bex). If they do, it prints a message saying that the modification has already been applied, and no changes are made. If they don't exist, it proceeds with the modifications. This way, running the code multiple times won't cause redundant changes.

- **24 values** were inserted in **Distance** to replace the NA's where the distance could be found by looking at the previous rows. The **46 remaining NA's** were then **removed** from Distance **leaving 0 remaining NA in the variable "Distance"**

2.10 Cleaning Female and Male ID

- Before cleaning Female and Male ID, here is a list of every dyad of the box experiment and their respective groups. This will help us find the missing names when only one individual is missing out of the duo (either male or female):
    a. Sirk & Sey - BD
    b. Ouli & Xin - BD
    c. Piep & Xia - BD
    d. Oerw & Nge - BD
    e. Oort & Kom - BD
    f. Ginq & Sho - AK
    g. Ndaw & Buk - Ak
    h. Xian & Pom - AK
    i. Guat & Pom - Ak
- Note that the 4 letter codes correspond to the femaleID, the 3 letter codes to the males ID and the 2 letter codes to the group name of the monkeys
- I need to check where are the NA's in both FemaleID and Male ID by looking at the rows where data is missing. Since every trial was made with a Dyad and never with an single individual, treating these two columns together makes more sense. If both individuals are missing I may have to delete the row.

```
## Row numbers with missing values in FemaleID:  865 866 867 868 869 870 871
872 873 874 875 876 877 878 879 1693 1694 1695 1696 1697 1698 1699 1700 1701
1702 1703 1704 1705 1706 1707 1708 1709 1710 1808 1809 1810 1811 1812 1813
1814 1815 1816 1817 1818 1819 1820 1884 1885 2619 2620 2621 2622 2623 2624
2625 2626 2627 2628 2629

## Number of missing values in FemaleID:  59

## Row numbers with missing values in MaleID:  1693 1694 1695 1696 1697 1698
1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710

## Number of missing values in MaleID:  18

## Number of rows with missing values in both FemaleID and MaleID:  18

## Row numbers with missing values in both FemaleID and MaleID:  1693, 1694,
1695, 1696, 1697, 1698, 1699, 1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707,
1708, 1709, 1710

## Number of missing values in FemaleID not in MaleID:  41

## Row numbers with missing values in FemaleID not in MaleID:  865, 866, 867,
868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 1808, 1809, 1810,
1811, 1812, 1813, 1814, 1815, 1816, 1817, 1818, 1819, 1820, 1884, 1885, 2619,
2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629
```

- **FemaleID** has **41 NA's** while they are **18 NA's** in **Male ID**
- In these missing data, we have **18 NA's** that are in common between FemaleID and MaleID which represents the totality of the missing values in MaleID
- All the missing data in MaleID are found in consecutive rows, from row **1693** to row **1710** and are from the group Noha (NH) on the 19th of april 2023. We can also see that trials had bee made in the same day, and looking at the time of the experiment, the previous trials made and the audience we can see that these NA's in female and male ID we can asses that the individuals involved were **Xian** for the female ID and **Pom** for the **MaleID**. I will thus replace these values using a condtion. These NA's in Noha (Trial 1693 to 1710) are the only NA's that MaleID has and are the only NA's of female ID in Noha. I will thus replace every NA of **MaleID NA in Noha** with **Pom** and every **Female ID NA in Noha** with **Xian**

```
## Number of remaining NA values in MaleID after replacement:  0

## Number of remaining NA values in FemaleID after replacement:  41

## Number of rows with missing values in both MaleID and FemaleID after
replacement:  0
```

- In order to clean FemaleID, I will use the data from the now complete MaleID. I will use conditions stating that depending which name is found in MaleID when there is an NA in FemaleID, a certain name will have to replace the NA in female ID

- Before automating the process I will check manually the data to see if they are any exceptions or mistakes

```
## Rows with missing values in FemaleID:

## # A tibble: 41 × 16
##    Time     Date                Group   MaleID FemaleID FemaleCorn
DyadDistance
##    <chr>    <dttm>              <chr>   <chr>  <chr>         <dbl>
<dbl>
##  1 09:31:55 2023-07-22 00:00:00 Ankhase Buk    <NA>              7
1
##  2 09:33:14 2023-07-22 00:00:00 Ankhase Buk    <NA>              7
1
##  3 09:34:07 2023-07-22 00:00:00 Ankhase Buk    <NA>              7
0
##  4 09:34:51 2023-07-22 00:00:00 Ankhase Buk    <NA>              7
0
##  5 09:36:59 2023-07-22 00:00:00 Ankhase Buk    <NA>              7
0
##  6 09:38:13 2023-07-22 00:00:00 Ankhase Buk    <NA>              7
1
##  7 09:39:26 2023-07-22 00:00:00 Ankhase Buk    <NA>              7
0
##  8 09:41:11 2023-07-22 00:00:00 Ankhase Buk    <NA>              0
0
##  9 09:42:17 2023-07-22 00:00:00 Ankhase Buk    <NA>              0
0
## 10 09:44:06 2023-07-22 00:00:00 Ankhase Buk    <NA>              0
1
## # ℹ 31 more rows
## # ℹ 9 more variables: DyadResponse <chr>, OtherResponse <chr>, Audience
<chr>,
##   IDIndividual1 <chr>, IntruderID <chr>, Remarks <chr>, MaleCorn <dbl>,
##   Intrusion <dbl>, AmountAudience <dbl>
```

If there is NA in femaleID, we will replace the value with - Sirk if MaleID is Sey - Ouli if MaleID is Xin - Piep if MaleID is Xia - Oerw if MaleID is Nge - Oort if MaleID is Kom - Ginq if MaleID is Sho - Ndaw if MaleID is Buk

```
library(dplyr)

# Using dplyr to summarize combinations of MaleID and FemaleID
combinations_summary <- Bex %>%
  group_by(MaleID, FemaleID) %>%
  summarise(Count = n(), .groups = 'drop') %>%
  arrange(desc(Count))

# View the results
print(combinations_summary)
```

```
## # A tibble: 20 × 3
##    MaleID FemaleID Count
##    <chr>  <chr>    <int>
##  1 Xia    Piep       576
##  2 Sey    Sirk       557
##  3 Kom    Oort       338
##  4 Sho    Ginq       278
##  5 Pom    Xian       259
##  6 Buk    Ndaw       245
##  7 Xin    Ouli       159
##  8 Nge    Oerw       153
##  9 Piep   Xia         35
## 10 Oort   Kom         29
## 11 Ouli   Xin         27
## 12 Oerw   Nge         19
## 13 Sirk   Sey         17
## 14 Buk    <NA>        15
## 15 Sey    <NA>        13
## 16 Nge    <NA>        11
## 17 Buk    Ginq         6
## 18 Pom    Guat         5
## 19 Xin    Oort         4
## 20 Kom    <NA>         2

## Number of NA values in MaleID:  0

## Number of NA values in FemaleID:  0
```

- After the use of the conditions in FemaleID I could see the changes where successfully done and that 0 NA's are remaining in both FemaleID and MaleID

2.12 Dyad Response (7)

- The last variable I still have to treat for NA's is DyadResponse. Before treating the NA's we had 47 NA's we know that we treated most of them we have only 7 remaining. These NA's can be found at the rows **871, 1163, 1219, 1339, 1579,1888 and 1962**

```
## Rows with missing values in DyadResponse:  871, 1163, 1219, 1339, 1579,
1888, 1962

## Lines with missing values in DyadResponse:

## # A tibble: 7 × 16
##   Time     Date                Group     MaleID FemaleID FemaleCorn
DyadDistance
##   <chr>    <dttm>              <chr>     <chr>  <chr>         <dbl>
<dbl>
## 1 09:39:26 2023-07-22 00:00:00 Ankhase   Buk    Ndaw              7
0
## 2 10:13:56 2023-06-23 00:00:00 Ankhase   Sho    Ginq              0
4
```

```
## 3 08:34:45 2023-06-17 00:00:00 Baie Dan… Kom    Oort              3
2
## 4 08:54:12 2023-06-09 00:00:00 Baie Dan… Xia    Piep              1
0
## 5 13:35:08 2023-05-03 00:00:00 Baie Dan… Kom    Oort              5
3
## 6 13:27:30 2023-01-18 00:00:00 Ankhase   Buk    Ndaw              5
4
## 7 08:36:49 2022-12-13 00:00:00 Baie Dan… Kom    Oort              8
4
## # ℹ 9 more variables: DyadResponse <chr>, OtherResponse <chr>, Audience
<chr>,
## #   IDIndividual1 <chr>, IntruderID <chr>, Remarks <chr>, MaleCorn <dbl>,
## #   Intrusion <dbl>, AmountAudience <dbl>
```

- Row 871: The previous row was tolerance at 1m and the next tolerance at 0 which means that the row 871 should be **Tolerance** for DyadResponse
- Row 1163: The value can not be found from the other rows so I will **delete** row 1163
- Row 1219: The previous row was not approaching at 2m and the next is tolerance at 2m and tolerance at 1m, which means that the row 1219 should be **Tolerance** for DyadResponse
- Row 1339: The previous row was tolerance at 0m while the next one was tolerance at 0m, which means that the tow 1339 should be **Tolerance** for DyadResponse
- Row 1579: The value can not be found from the other rows so I will **delete** row 1579
- Row 1888: The value can not be found from the other rows so I will **delete** row 1888
- Row 1962: The value can not be found from the other rows so I will **delete** row 1962

```
## Number of remaining NA values in DyadResponse:  0
```

2.13 Final check: remaing NA's in Bex?

```
## Final check of NA values in Bex:

##           Time           Date          Group         MaleID       FemaleID
##              0              0              0              0              0
##     FemaleCorn   DyadDistance   DyadResponse  OtherResponse       Audience
##              0              0              0              0              0
##  IDIndividual1     IntruderID        Remarks       MaleCorn      Intrusion
##              0              0              0              0              0
## AmountAudience
##              0
```

3. Correction and creation of New Variables

3.1 Making a backup of Bex

- I will now continue making changes, in case they are any problems or comparason to be made I may use this function to look how Bex looked at that point

3.2 Treating Remarks before processing with new modifications

- Since I have removed all the missing data from the different columns, I now have to correct potential mistakes that can be found and create new variables to be able to manipulate better my data. Since the column remarks contains corrections and additional information, so i will treat it now

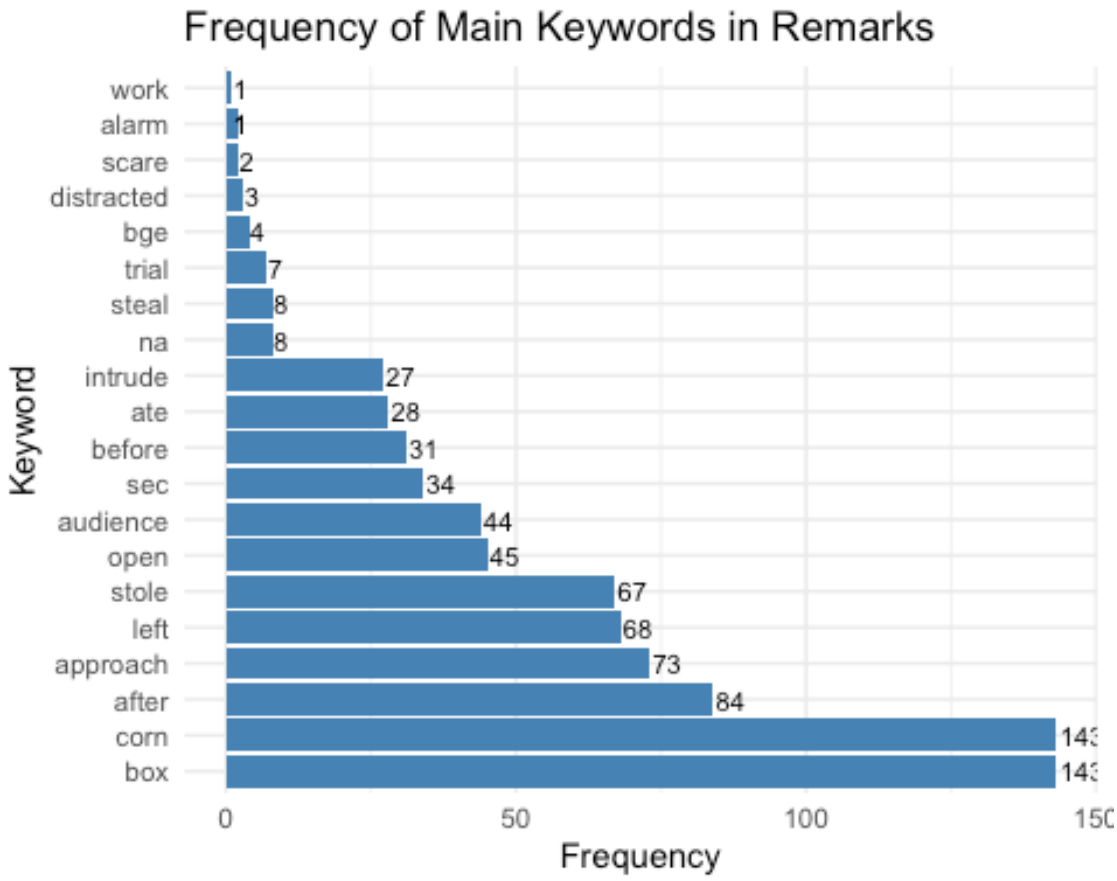
- Before that lets check how many remarks we have in our dataset, how many of the main keywords we can find and make a visual representation of it

3.2.2 Vizualization of the Remarks Keywords

- Before making any changes I will make a count of the total amount of remarks and a count and barplot of the main keywords in the column to see in which proportion they are found. It has to be noted that some of the words are used in different contexts and have different meaning. This is why I will clean them manually

```
## Number of 'No Remarks' entries:  2139

## Number of actual remarks entries:  603
```

- There will be 599 entries I will have to treat manually in the Excel Spreadsheet for the Remarks




```
## Total number of keyword occurrences in the Barplot:  822
```

3.2.3 Exporting of Bex

- I will now export the dataset and treat manually treat the remarks in an Excel spreadsheet before uploading it again and creating a new dataframe. I will also print a Glimpse of Bex to have information before the manual changes

```
## Glimpse of the Bex Before treating Remarks:

## Rows: 2,742
## Columns: 16
## $ Time           <chr> "09:47:50", "09:50:07", "09:53:11", "09:54:28",
"09:55:…
## $ Date           <dttm> 2022-09-27, 2022-09-27, 2022-09-27, 2022-09-27,
2022-0…
## $ Group          <chr> "Baie Dankie", "Baie Dankie", "Baie Dankie", "Baie
Dank…
## $ MaleID         <chr> "Nge", "Nge", "Nge", "Nge", "Nge", "Nge", "Nge",
"Nge",…
## $ FemaleID       <chr> "Oerw", "Oerw", "Oerw", "Oerw", "Oerw", "Oerw",
"Oerw",…
## $ FemaleCorn     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 7,
7, 7…
## $ DyadDistance   <dbl> 2, 2, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 2, 1, 1,
```

```
0, 0…
## $ DyadResponse   <chr> "Tolerance", "Tolerance", "Tolerance", "Tolerance",
"To…
## $ OtherResponse  <chr> "No Response", "No Response", "No Response", "No
Respon…
## $ Audience       <chr> "Obse; Oup; Sirk", "Obse; Oup; Sirk", "Oup; Sirk",
"Sir…
## $ IDIndividual1  <chr> "No individual", "No individual", "No individual",
"No …
## $ IntruderID     <chr> "No Intrusion", "No Intrusion", "No Intrusion", "No
Int…
## $ Remarks        <chr> "No Remarks", "No Remarks", "Nge box did not open
becau…
## $ MaleCorn       <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
3, 3…
## $ Intrusion      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
0, 0…
## $ AmountAudience <dbl> 3, 3, 2, 1, 2, 2, 2, 1, 1, 2, 6, 6, 3, 2, 2, 2, 2,
2, 2…

## [1] "/Users/maki/Desktop/Master Thesis/BEX 2223 Master Thesis Maung
Kyaw/IVPToleranceBex2223"
```

3.2.4 Journal of manual changes in Bex excel spreadsheet

- Before treating all the data in the Remarks I will create a few columns to redistribute information

    1. **Context**: To add contextual information
    2. **SpecialBehaviour** : To report any particular behaviour an individual could have done during a trial
    3. **Got corn**, to see if the individual got the corn or not

- Also whenever i will have treated a remark, i will replace it with “Treated”. And if I have to delete the row I’ll write “Delete”. After re importing the data I will make a count of these changes to see if I still have the correct amount of cells and changes that have been done

- 1. Creation of the columns **Context**, **SpecialBehaviour** and **GotCorn**

- 2. Default values for the new columns are **NoContext**, **NoSpecialBehaviour** & **Yes**

  a.Context: **BoxMalfunction**, **BoxOpenedBefore**, **NoExperiment**, **Agonistic**, **Guat;Ap;Xian**, **CornLeak**, **BetweenGroupEncounter**, **ContactCalling**,

  b.SpecialBehaviour **Oerw;Vo;Exp**, **Sey;Ap;AfterOpen**, **Oerw;Vo;Exp,Nge;Vo;Exp**, **Sirk;ApAfter30**, **Sirk;Av;Oerw**, **Oerw;Lo,Sey;Sf;Oort,Oort;At;Kom**, **Kom;Ap;AfterOpen**, **Sey;Ch,Sirk**, **Xin;Hesitation**. **Xia;Sf;Piep**, **Pom;Sf;Xian**,**Kom;Sf;Oort**, **Sey;Sf;Sirk**, **Xia;Sf;Piep,Piep;Sf,XIa**, **Oort;At;Kom**,

Sey;Rt;Sho;Ap, **Sho;Rt;Ginq;Ap**, **Buk;Sf;Ndaw**, **Sho;Rt;Ndaw;Ap**, **Oort;Sf;Kom**, **Ginq;Sho;Ap;After30**, **Ndaw;Sc,Buk;Sf**, **Ndaw;Ap;After30**, **Kom;Ap;After30**, **Xia;Piep;Ap;After30**, **Pom;Bi;Xian**, **Sho;Ndaw;Av;Buk**, **Kom;Sf;Oort**, **Kom;St;Oort,Oort;St;Kom**, **Sey;Hi;Sirk**, **Obse;Ap;Piep;Av,Piep;Sf;Xia**, **Sirk;ApWhenPartnerLeft**, **Sey;Hh;Sirk**, **Xia;Sf;Piep;Sc**, **Xia;Piep;ShareFood**, **Piep;Ap;After30,Xia;Mu;Piep**, **Oort;St;Sirk;Ja,Sey;Sf**, **Pom;Sf;Xian**, **Ndaw;ApWhenPartnerLeft**, **Xian;At;Pom,Gaya;Su**, **Xian;Sf;Pom**, **Xian;Hesitation**, **Xia;ApWhenPartnerLeft,Sirk;Hesitation**, **Ginq;Hesitation**, **Sey;Ap;Kom;Av**, **Oort;Sc;Kom**, **Xian; Pom**, **Pom;Ap;Xian**, **Pom;Ap;Xian,Xian;Rt**, **Sey;Ap;Sirk;Rt**, **Sey;St;Sirk;Ig**, **Xia;Asf;Piep**, **Piep;ApWhenPartnerLeft**, **Sho;Ap;After30**, **Ginq;ApWhenPartnerLeft**, **Pom;Sf;Xian;Sf;Pom**, **Xian;ApWhenPartnerLeft**, **Piep;Ch;Sirk**, **Sey;St;Sirk**, **Ndaw;Ap;After30**, **Xian;Ap;After30**, **Xian;St;Prai**, **Pom;Sf;Xian;Vc**, **Kom;Ap;After30**, **Kom;ApproachWithPartner**, **Oort;ApWhenPartnerLeft**, **Sho;Ap;After30,Ginq;Ap;After30**, **Ginq;ApproachWithPartner**, **Ndaw;Hesitation**, **Oerw;Hesitation**, **Oerw;ApWhenPartnerLeft**. **Piep;Ap;After30**, **Sirk;Ap;After30**, **Xia;Ap;After30**, **Ouli;Gr;BBOuli**, **Oerw;Ap;After30**, **Sirk;Hesitation**, **Sey;Ap;Sirk;Av**, **Ouli;Ap;Xia;Av**, **Xin;Ap;After30**, **Sho;Sf;Ginq;Sc**, **Xia;ApWhenPartnerLeft**, **Sey;Ap;Sirk;Ja**, **Nge;Oerw;ShareFood**, **Nge;Ap;Oerw;Oerw;At,Obse;At;Nge**,

c.GotCorn: **No;Nge**, **No;Piep**, **No;Xian**, **No;Oort**, **No;Sirk**, **No;Kom**, **No;Ndaw**, **No;Kom**, **No;Oort**, **No;Xia**, **No;Buk**, **No;Sho**, **No;Sey,No;Piep**, **No;Ginq**

d. Remarks: **Treated**, **TODelete**

-

4. Values set for exsting columns

a. IntruderID: **Sey**, **Oerw**, **Guat**, **Kom**, **Gris**, **Sho**, **Oerw; Ouli**, **Guat; Gri**, **Xop**, **Obse**, **Oort**, **Obse; Sey**, **Ginq; Ghid**, **Xia**, **Grif**, **Sey**, **Gree; Gran**, **Godu; Gub**, **Gran**, **Oerw; Nak**, **Ghid**, **Buk**, **Oup**

b. DyadDistance: **6**, **7**, **8**, **9** , **1**

c. Audience: **UnidentifiedAudience**, **Ouli; Riss**, **Gris**, **Sey**, **Sey; Piep; Sirk, Oup Ome**

d. IDIndividual1: **Piep**, **Oort; Kom**, **Ndaw; Buk**, **Sho; Ginq**, **Ndaw, Buk**, **Xian, Pom**, **Oort; Kom**, **Buk; Ndaw**, **Sirk; Sey**, **Xin; Ouli**, **Oerw; Nge**

e. DyadResponse: **Tolerance**, **Not approaching; Losing interest**, **Losing interest; Intrusion**

UPDATES FROM HERE

3.3 Re Uploading the dataset after the treatment of the Remarks

- Now that I have … I will reimport the dataset and make a check of the data

```
## Rows: 2,742
## Columns: 16
## $ Time           <chr> "09:47:50", "09:50:07", "09:53:11", "09:54:28",
"09:55:…
## $ Date           <dttm> 2022-09-27, 2022-09-27, 2022-09-27, 2022-09-27,
2022-0…
## $ Group          <chr> "Baie Dankie", "Baie Dankie", "Baie Dankie", "Baie
Dank…
## $ MaleID         <chr> "Nge", "Nge", "Nge", "Nge", "Nge", "Nge", "Nge",
"Nge",…
## $ FemaleID       <chr> "Oerw", "Oerw", "Oerw", "Oerw", "Oerw", "Oerw",
"Oerw",…
## $ FemaleCorn     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 7,
7, 7…
## $ DyadDistance   <dbl> 2, 2, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 2, 1, 1,
0, 0…
## $ DyadResponse   <chr> "Tolerance", "Tolerance", "Tolerance", "Tolerance",
"To…
## $ OtherResponse  <chr> "No Response", "No Response", "No Response", "No
Respon…
## $ Audience       <chr> "Obse; Oup; Sirk", "Obse; Oup; Sirk", "Oup; Sirk",
"Sir…
## $ IDIndividual1  <chr> "No individual", "No individual", "No individual",
"No …
## $ IntruderID     <chr> "No Intrusion", "No Intrusion", "No Intrusion", "No
Int…
## $ Remarks        <chr> "No Remarks", "No Remarks", "Nge box did not open
becau…
## $ MaleCorn       <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
3, 3…
## $ Intrusion      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
0, 0…
## $ AmountAudience <dbl> 3, 3, 2, 1, 2, 2, 2, 1, 1, 2, 6, 6, 3, 2, 2, 2, 2,
2, 2…

## Rows: 2,742
## Columns: 19
## $ Time             <chr> "09:47:50", "09:50:07", "09:53:11", "09:54:28",
"09:5…
## $ Date             <dttm> 2022-09-27, 2022-09-27, 2022-09-27, 2022-09-27,
2022…
## $ Group            <chr> "Baie Dankie", "Baie Dankie", "Baie Dankie",
"Baie Da…
## $ MaleID           <chr> "Nge", "Nge", "Nge", "Nge", "Nge", "Nge", "Nge",
"Nge…
## $ FemaleID         <chr> "Oerw", "Oerw", "Oerw", "Oerw", "Oerw", "Oerw",
"Oerw…
## $ FemaleCorn       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7,
7, 7,…
## $ DyadDistance     <dbl> 2, 2, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 2, 1,
```

```
1, 0,…
## $ DyadResponse     <chr> "Tolerance", "Tolerance", "Tolerance",
"Tolerance", "…
## $ OtherResponse    <chr> "No Response", "No Response", "No Response", "No
Resp…
## $ Audience         <chr> "Obse; Oup; Sirk", "Obse; Oup; Sirk", "Oup;
Sirk", "S…
## $ IDIndividual1    <chr> "No individual", "No individual", "No
individual", "N…
## $ IntruderID       <chr> "No Intrusion", "No Intrusion", "No Intrusion",
"No I…
## $ Remarks          <chr> "No Remarks", "No Remarks", "Treated", "Treated",
"No…
## $ MaleCorn         <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
3, 3,…
## $ Intrusion        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0, 0,…
## $ AmountAudience   <dbl> 3, 3, 2, 1, 2, 2, 2, 1, 1, 2, 6, 6, 3, 2, 2, 2,
2, 2,…
## $ Context          <chr> "NoContext", "NoContext", "BoxMalfunction",
"BoxOpene…
## $ SpecialBehaviour <chr> "NoSpecialBehaviour", "NoSpecialBehaviour",
"Oerw;Vo;…
## $ GotCorn          <chr> "Yes", "Yes", "No;Nge", "Yes", "Yes", "Yes",
"Yes", "…
```

- I will now check the state fo remarks and of the data before and after to see if no major mistakes may have been done
- First the amount of Remarks and No Remarks, before: Number of 'No Remarks' entries: 2119 Number of actual remarks entries: 599

```
## Number of 'No Remarks' entries:  2135

## Number of actual remarks entries:  607

## Total number of keyword occurrences in the Remarks column:  0
```

- NA Check in BexClean

```
## Number of NA entries in Context:  0

## Number of NA entries in SpecialBehaviour:  0

## Number of NA entries in GotCorn:  0

## Number of NA entries in BexClean:  0
```

###4.1 Creation of new dataframe?

Treatment and cleaning of last variables before stats!?

4.2 Time - Creation of Period and Hour

- Time : I did not plan to use this variable but since I used, I considered looking at the time sections in which we did the expermiment. I will thus look at the time ranges (max and min in the day / latest and earliest time) before separating the day in different sections to have an idea in which part of the day most of the experiments occured. This will not be used in my analysis, but if I wanted to, I could interesting to compare the amount of experimentations made per day and have a line indicating the time of sunrise.

- The **Minimum Time** in the dataset is **06:03:26*** while the **Maximum Time** is at **16:36:59**

- In my box experiment I have this variable called time that tells me when the experiment was done. I don't think I need this information per se. I was wondering if it could be easy and interesting to see from when to when the time occurs and then separate this time in a few sections like early, monring, morning, miday, afternoon, end of the day

- a.6 to 8 : Early morning b.8 to 10: Morning c.10 to 12: Noon d.12 to 14: Afternoon e.14 to 17: End of the day

- Last, I want to create a variable called Hour that will take the value in Time and round it to the hour in which it is ex: from 06:00 to 06:59 -> 6, from 07:00 to 07:59 -> 7 etc…

- This will allow me to see when most of the trials occured with more detail and I will be to see in which hour most of the trial happened. Nevertheless Period will be better for an improved readability

Date - Creation of Month and Day

- i want to Create a variable called month to see the month of the experiment and day so I know which day of the experiment it was (1st, 10th, 1000th..)

Group - Ok

Male and Female ID - Creation of Dyad, Trial and Session

- I will use Female and Male ID to create different variables
    1. While checking if there are still any mistakes in **FemaleID and MaleID** using **unique**, I saw that some of the names are in the wrong rows. I want the **3 letter male codes** whether they are in the column FemaleID or MaleID to be in the **new column Male** while I want the **4 letter female codes** whether they are in FemaleID or MaleID to be in the **new column Female** before checking again with unique that the transformation worked. I will use mutate

```
## Unique Female IDs: Sirk Ginq Piep Oerw Xin Ndaw Xia Sey Ouli Nge Oort Xian
Guat Kom
```

```
##
## Ginq Guat  Kom Ndaw  Nge Oerw Oort Ouli Piep  Sey Sirk  Xia Xian  Xin
##  283    5   29  259   19  164  341  159  575   17  570   35  259   27

## Unique Male IDs: Sey Sho Xia Nge Ouli Buk Piep Sirk Xin Oerw Kom Pom Oort

##
##  Buk  Kom  Nge Oerw Oort Ouli Piep  Pom  Sey  Sho Sirk  Xia  Xin
##  265  337  164   19   29   27   35  264  570  277   17  575  163

##
##        Buk Kom Nge Oerw Oort Ouli Piep Pom Sey Sho Sirk Xia Xin
##   Ginq   6   0   0    0    0    0    0   0   0 277    0   0   0
##   Guat   0   0   0    0    0    0    0   5   0   0    0   0   0
##   Kom    0   0   0    0   29    0    0   0   0   0    0   0   0
##   Ndaw 259   0   0    0    0    0    0   0   0   0    0   0   0
##   Nge    0   0   0   19    0    0    0   0   0   0    0   0   0
##   Oerw   0   0 164    0    0    0    0   0   0   0    0   0   0
##   Oort   0 337   0    0    0    0    0   0   0   0    0   0   4
##   Ouli   0   0   0    0    0    0    0   0   0   0    0   0 159
##   Piep   0   0   0    0    0    0    0   0   0   0    0 575   0
##   Sey    0   0   0    0    0    0    0   0   0   0   17   0   0
##   Sirk   0   0   0    0    0    0    0   0 570   0    0   0   0
##   Xia    0   0   0    0    0    0   35   0   0   0    0   0   0
##   Xian   0   0   0    0    0    0    0 259   0   0    0   0   0
##   Xin    0   0   0    0    0   27    0   0   0   0    0   0   0
```

	Buk	Kom	Nge	Oerw	Oort	Ouli	Piep	Pom	Sey	Sho	Sirk	Xia	Xin
Ginq	6	0	0	0	0	0	0	0	0	277	0	0	0
Guat	0	0	0	0	0	0	0	5	0	0	0	0	0
Kom	0	0	0	0	29	0	0	0	0	0	0	0	0
Ndaw	259	0	0	0	0	0	0	0	0	0	0	0	0
Nge	0	0	0	19	0	0	0	0	0	0	0	0	0
Oerw	0	0	164	0	0	0	0	0	0	0	0	0	0
Oort	0	337	0	0	0	0	0	0	0	0	0	0	4
Ouli	0	0	0	0	0	0	0	0	0	0	0	0	159
Piep	0	0	0	0	0	0	0	0	0	0	0	575	0
Sey	0	0	0	0	0	0	0	0	0	0	17	0	0
Sirk	0	0	0	0	0	0	0	0	570	0	0	0	0
Xia	0	0	0	0	0	0	35	0	0	0	0	0	0
Xian	0	0	0	0	0	0	0	259	0	0	0	0	0
Xin	0	0	0	0	0	27	0	0	0	0	0	0	0

1. Create a variable called Male that in each row will take the name of the **3 letter code** that is either in MaleID or Female ID and a variable called Female that in each row will take the name of the **4 letter code** that is either in MaleID or FemaleID

```
##  [1] "Sey Sirk" "Sho Ginq" "Xia Piep" "Nge Oerw" "Xin Ouli" "Buk Ndaw"
##  [7] "Buk Ginq" "Kom Oort" "Pom Xian" "Pom Guat" "Xin Oort"
```

Var1	Freq
Buk Ginq	6
Buk Ndaw	259
Kom Oort	366
Nge Oerw	183
Pom Guat	5
Pom Xian	259
Sey Sirk	587
Sho Ginq	277
Xia Piep	610
Xin Oort	4
Xin Ouli	186

Correct from here, mistake were inserted in dyad by changin wrong lines

2. Create the variable called **Dyad** created by combining the name of FemaleID and MaleID into one name with a space between the two codes. For information the 3 letter code is the name of the female while the 4 letter code is the name of the male like displayed here;

New output! : 2673 trials Old output : 2724 trials

–>51 rows missing!

- Buk Ginq 6
- Buk Ndaw 257 vs 255!
- Kom Oort 366 vs 324! +26!
- Kom Ginq! 2!
- Nge Oerw 181 vs 163 + 18!
- Pom Guat 5
- Pom Xian 257 vs 251!
- Sey Sirk 584 vs 544! vs 17!
- Sho Ginq 273 vs 272!
- Xia Piep 606 vs 35!+ 567!
- Xin Oort 4

- Xin Ouli 185 vs 27!+ 157!
  - They are a few **wrong dyads** that I will have to identify in the dataset and manuallly correct, those wrong dyads to change and identify are: -Buk Ginq - 6 occurences -Xin Oort - 4 occurences

```
##  [1]  613  614  615  616  617  931 2710 2711 2712 2713

## [1] "Buk Ginq" "Xin Oort"
```

-
  - I will change the occurences of **Buk Ginq** to **Sho Ginq** for **row 605 to 609** and **row 921**. I know these triasl are with SHo Ginq because the comments metionned Sho in them while Male(ID) gave Buk
  - For the **rows from 2692 to 2695** since, Ouli is in the audience it is unlikely that we had trials with the dyad **Xin Ouli**. Also I think they are little chances that the names of both individuals were entered wrong. I will replace these occurences where we had **Xin Oort** by **Kom Oort**
  - I thus want **Buk** to be replaced in male ID in rows 605 to 609 and row 921 with **Sho** and, **Xin** to be replaced by **Kom** in rows 2692 to 2695 in Male ID before updating Dyad

```
## Unique Male after correction Sey Sho Xia Nge Ouli Buk Piep Sirk Xin Oerw
Kom Pom Oort

## Unique Dyad after correction Sey   Sirk Sho   Ginq Xia   Piep Nge   Oerw
Xin   Ouli Buk   Ndaw Sho   Ndaw Buk   Ginq Kom   Oort Sho   Oerw Pom   Xian
Pom   Guat Kom   Ginq Kom   Ouli Xin   Oort
```

Var1	Freq
Buk	263
Kom	341
Nge	163
Oerw	19
Oort	29
Ouli	27
Piep	35
Pom	264
Sey	570
Sho	277
Sirk	17
Xia	575
Xin	162

Var1	Freq
Buk Ginq	6
Buk Ndaw	257
Kom Ginq	3
Kom Oort	366
Kom Ouli	1
Nge Oerw	182
Pom Guat	5
Pom Xian	259
Sey Sirk	587
Sho Ginq	274
Sho Ndaw	2
Sho Oerw	1
Xia Piep	610
Xin Oort	4
Xin Ouli	185

3. Create the variable called **Trial** where the data will be **sorted by date and dyad** in order to see how many trials have been done with each individual: **One row (per dyad) = one trial** and the variable called **Day** where the data will be **sorted by date and dyad and day** in order to see how many sessions have been done with each individual: **One day (per dyad) = one session** Now, let's proceed with creating the Dyad variable, Trial, and Day:

4. Make a summary of trial and session so I can see see how many trials and sessions have been done with the individuals

```
## Trial Summary:

## # A tibble: 15 × 2
##    Dyad       Num_Trials
##    <chr>           <int>
##  1 Buk   Ginq          6
##  2 Buk   Ndaw        257
##  3 Kom   Ginq          3
##  4 Kom   Oort        366
##  5 Kom   Ouli          1
##  6 Nge   Oerw        182
##  7 Pom   Guat          5
##  8 Pom   Xian        259
##  9 Sey   Sirk        587
## 10 Sho   Ginq        274
## 11 Sho   Ndaw          2
## 12 Sho   Oerw          1
## 13 Xia   Piep        610
```

```
## 14 Xin   Oort          4
## 15 Xin   Ouli        185

##
## Day Summary:

## # A tibble: 15 × 2
##    Dyad       Num_y
##    <chr>      <int>
##  1 Buk   Ginq     2
##  2 Buk   Ndaw    41
##  3 Kom   Ginq    42
##  4 Kom   Oort    73
##  5 Kom   Ouli    74
##  6 Nge   Oerw    96
##  7 Pom   Guat    97
##  8 Pom   Xian   116
##  9 Sey   Sirk   169
## 10 Sho   Ginq   204
## 11 Sho   Ndaw   205
## 12 Sho   Oerw   206
## 13 Xia   Piep   255
## 14 Xin   Oort   256
## 15 Xin   Ouli   283
```

Var1	Freq
Buk Ginq	6
Buk Ndaw	257
Kom Ginq	3
Kom Oort	366
Kom Ouli	1
Nge Oerw	182
Pom Guat	5
Pom Xian	259
Sey Sirk	587
Sho Ginq	274
Sho Ndaw	2
Sho Oerw	1
Xia Piep	610
Xin Oort	4
Xin Ouli	185

5. After relfection I decided to **remove every column** that is with PomGuat since they are not enough trials for this Dyad and since we then changed PomGuat for PomXian.

I have 5 occurnces to change. For easier manipulation I will remove every row where there is **Guat**

```
## Change in Rows:  -5
```

Female Corn and Male Corn

- The idea is that for each Dyad, we gave an amount of corn to attract the monkey of a dyad to the right distance of his partner for a trial by putting corn in experiment box that he could get by approaching. We repeated this step as much as needed to have our Dyad at the desired distance to continue the trials from the previous day of experimentation. This means I will only Keep the last number per dyad and day for each day.
- I decided to create **two variables**, that are called **PlacementMale** and **PlacementFemale** that will only keep the final amount of corn given to each individual within a day of experiment

DyadDistance - Creation of proximity

- I would like to create a variable called proximity to have another measure of the proximity of the individuals.
- First lets look at the maximum and minimum distance found in Dyad Distance

```
## Maximum Distance: 10

## Minimum Distance: 0
```

- The **minumum Distance** is **0m** while the **maximum Distance** is **10m**
- Now lets create a new variable called **Proximity** using the distances found in **DyadDistance** on the following model:
    a. 0 = Contact
    b. 1 - 2 = Very close
    c. 2 - 3 = Close
    d. 4 - 5 = Distant
    e. 5 - 6 = Far
    f. 7 - 8 = Very Far
    g. 9 - 10 = Maximum Distance
- The idea is that for each Dyad, we gave an amoun of corn to attract the monkey of a dyad to the right distance of his partner for a trial by putting corn in experiment box that he could get by approaching. We repeated this step as much as needed to have our Dyad at the desired distance to continue the trials from the previous day of experimentation. This means I will only Keep the last number per dyad and day for each day
- DyadDistance - Creation of proximity
- Dyad Response - Detailed cleaning

- Reminder: The different behaviors that are coded in **DyadResponse** are: **Distracted**, **Female aggress male**, **Male aggress female**, **Intrusion**, **Loosing interest**, **Not approaching**, **Tolerance** and **Other**

  - I will change the columns associated to each behavior (i.e. Response) of **DyadResponse** into dichotomic variables in order to see the frequency of each behaviour
  - This will allow me to see which behavior occurred more than others, and what differences are between dyads
  - As multiple response could occur within the same trial, multiple behaviors can be found in a single cell. I will create a hierarchy to reduce the amount of behaviors assigned to each trial (if there is more than one). This will also be complemented with the information found in the column comments
    1. correct any mistakes (ex. if tolerance and aggression are together aggression>tolerance)
    2. assign as few labels per trial
    3. get a better View and understanding of the data and the most common behaviours produced by each dyad
    4. create variables that can complement the behaviour found (ex. not approaching + looks at partner would be looks at partner + a new variable called hesistant to see when the did not come but look at the other individual /)
    - Create a table with each combination existing
    - Decide what is more important

- **Dyad Response Hierarchy** Projection of the hierarchy (changes will be made) - Aggression > Tolerance - Tolerance > Not approaching -> Create a variable called hesistant in addtion to the tolerance count to see frequency of tolerance behaviour that happened after > 1min - Tolerance > Loosing interest - Tolerance > Intrusion - Not approaching = looking box but not coming while Loosing interest = not paying attention to the box - Intrusion > Loosing interest - Intrusion > Not approaching - Not approaching > Looks at partner - We can code every look at partner as no approaching and keep the count of looks at partner as additional information - Not approaching >?> Loosing interest ? !! - Define distracted - Not approaching > Distracted - Aggression > Not approaching - Other > Look case by case and categorize depending of behavior
  - Remarks may be used for the same reason

  - First I want to see how many rows in DyadResponse have more than one entry per cell

    - And **if there are more than one value per cell**, report it in a new column called “MultipleResponses”

```
## Number of rows with multiple entries in DyadResponse:  233
```

```
## Rows with multiple entries in DyadResponse:  57, 62, 68, 72, 75, 88, 92,
93, 101, 102, 117, 137, 196, 198, 248, 282, 284, 285, 296, 305, 309, 311,
325, 329, 333, 334, 335, 348, 353, 354, 374, 375, 394, 395, 398, 404, 416,
461, 475, 476, 489, 504, 517, 518, 529, 536, 543, 604, 607, 629, 632, 744,
752, 753, 761, 764, 765, 767, 772, 773, 784, 790, 821, 832, 864, 869, 891,
895, 901, 915, 922, 937, 940, 941, 957, 958, 965, 976, 1000, 1007, 1018,
1030, 1045, 1051, 1063, 1212, 1216, 1228, 1232, 1243, 1246, 1248, 1250, 1257,
1274, 1287, 1291, 1296, 1311, 1313, 1314, 1329, 1330, 1333, 1339, 1344, 1346,
1355, 1381, 1401, 1402, 1410, 1419, 1464, 1467, 1476, 1481, 1487, 1491, 1492,
1498, 1513, 1523, 1524, 1531, 1561, 1598, 1601, 1605, 1610, 1619, 1632, 1633,
1636, 1639, 1642, 1645, 1648, 1657, 1661, 1717, 1721, 1727, 1739, 1748, 1750,
1782, 1783, 1786, 1793, 1798, 1799, 1801, 1802, 1803, 1804, 1805, 1812, 1813,
1816, 1817, 1818, 1819, 1822, 1825, 1826, 1827, 1828, 1829, 1834, 1840, 1853,
1854, 1855, 1886, 1894, 2085, 2087, 2088, 2098, 2099, 2100, 2101, 2102, 2106,
2110, 2144, 2145, 2146, 2149, 2150, 2163, 2172, 2181, 2182, 2189, 2190, 2191,
2196, 2198, 2199, 2202, 2215, 2223, 2234, 2245, 2254, 2287, 2291, 2340, 2350,
2351, 2356, 2359, 2360, 2367, 2392, 2393, 2397, 2462, 2463, 2464, 2490, 2531,
2602, 2610, 2627, 2630, 2631, 2644, 2652, 2658, 2730
```

- Now that I know that they are 230 rows with multiple entries I will print each combinations to see which one I have to print and also display the combinations once the data is split.
- I will make an intermediary summary to see that state of DyadResponse at the moment
- A. Summary od DyadResponse
- B.Idnetify Rows with more than 1 entry
- Step 1: Identify Rows with More than 1 Entry
  beforesummaryrowswithmultipleentries_1 <- which(sapply(Bex$DyadResponse, function(x) length(unlist(strsplit(as.character(x), ";"))) > 1))

Display the rows with more than 1 entry in DyadResponse
knitr::kable(Bex[beforesummaryrowswithmultipleentries_1,])

- Print the amount of cells with more than 1 entry and the total number of rows
  cat("Number of rows with more than 1 entry in DyadResponse:", length(beforesummaryrowswithmultipleentries_1), "")
- C. Identify Rows with more than 2 entry

```
## Number of rows with more than 1 entry in DyadResponse:  233
```

-

```
## Unique Combinations and Counts for More than 1 Entry:

##  Distracted &  Intrusion  2
##  Female aggress male &  Intrusion 2
##  Looks at partner &  Other 1
```

```
##  Losing interest &  Intrusion 1
##  Losing interest &  Looks at partner 1
##  Male aggress female &  Female aggress male 2
##  Male aggress female &  Looks at partner 1
##  Not approaching &  Intrusion 1
##  Not approaching &  Losing interest 1
## Distracted &  Losing interest 1
## Female aggress male &  Intrusion 1
## Female aggress male &  Not approaching 3
## Male aggress female &  Intrusion 1
## Male aggress female &  Looks at partner 1
## Male aggress female &  Not approaching 2
## Not approaching &  Distracted 7
## Not approaching &  Intrusion 21
## Not approaching &  Looks at partner 10
## Not approaching &  Losing interest 46
## Not approaching &  Other  1
## Tolerance &  Distracted   3
## Tolerance &  Female aggress male 23
## Tolerance &  Intrusion    42
## Tolerance &  Looks at partner 25
## Tolerance &  Losing interest 13
## Tolerance &  Male aggress female 31
## Tolerance &  Not approaching 6
## Tolerance &  Other        8

## Change in Occurrences for the Chunk (Female aggress male > Not
approaching):  2731 2737 2737 2478 2737 2737 2737 2737 2737 2737 2737 2737
2737 2737 2737 2737 2737 2737 2737 2737 2400 2737 2737 2737 2737 2737 2737
2737 2737 2737 2737 2737 2737 2573 2737 2737 2737 2737 2737 2737 2737 2737
2737 2737 2737 2737 2718 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737
2737 2708 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737
2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2710 2737 2737 2737
2737 2737 2737 2737 2737 2737 2737 2737 2702 2737 2737 2737 2732 2737 2737
2737 2737 2737 2737 2737 2737 2737 2737 2478 2737 2737 2737 2737 2737 2737
2737 2737 2737 2737 2737 2167 2737 2737 2737 2460 2737 2737 2737 2737 2737
2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737
2737 2737 2720 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737
2162 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2737 2733 2578 2737
2737 2737 2737 2737 2737
```

6. Male agress female > Not approaching : Remove Not approaching if there is Male aggress female (2x)

```
## Change in Occurrences for the Chunk (Male aggress female > Not
approaching):  -2
```

7. Tolerance > Distracted: Remove Distracted if there is tolerance (3x)

```
## Change in Occurrences for the Chunk Tolerance > Distracted:  -3
```

8.Female agress male > Tolerance: Remove Tolerance if there is Female agress male (23x)

```
## Change in Occurrences for the Chunk (Tolerance & Female aggress male):  -
23
```

9.Tolerance > Loosing interest: Remove Losing interest if there is Tolerance (13x)

```
## Change in Occurrences for the Chunk (Tolerance & Losing interest):  0
```

10.Male agress female > Tolerance : Remove Tolerance if there is Male agress male (31)

```
## Change in Occurrences for the Chunk (Tolerance & Male aggress female):  -
29
```

11.Tolerance > Not approaching: Remove Not approaching if there is tolerance (6x)

```
## Change in Occurrences for teh Chunk (Tolerance & Not approaching):  -6
```

- Looks at partner & Other 1 - Display line for detailed check (1x)
- Not approaching & Losing interest 46 - Keep both, I will consider doing a detailled check
- Not approaching & Distracted 7 - Display line for detailed check
- Not approaching & Other 1 - Display line for detailed check
- Tolerance & Other 8 - Display line for detailed check
- Code to display rows numbers and the response in DyadResponse that have mroe than one entry
- Other Reponse - DEtailed cleaning to delete the column
- Audience - Creation of Amount Audience and Density
- ID Individua1 - Not sure yet
- Intruder ID
- Remarks - Detailed Cleaning
- Intrusion
- Not Approaching
- Lossing Interest
- Distracted
- MultipleResponse
- Amount Audience
- DyadDistance

- Distance
- No trial

Code to keep and place again

Code to settle

I also chose to directly create new dichotomic variables for “Not approaching”, “Intrusion”, “Losing interest”, “Distracted”, for this I would like the function to 1.Check if there is a **value different than “No individual”** 2.**If the value ≠“No individual”** then I want it to **take the response found in “DyadResposne”**

```
## Occurrences of 'Not approaching' in DyadResponse: 560

## Occurrences of 'Intrusion' in DyadResponse: 116

## Occurrences of 'Losing interest' in DyadResponse: 97

## Occurrences of 'Distracted' in DyadResponse: 14

## Warning: Unknown or uninitialised column: `NotApproaching`.

## Warning: Unknown or uninitialised column: `LosingInterest`.

## Warning: Unknown or uninitialised column: `Distracted`.

## Occurrences of '1' in 'NotApproaching': 0

## Occurrences of '1' in 'Intrusion': 56

## Occurrences of '1' in 'LosingInterest': 0

## Occurrences of '1' in 'Distracted': 0
```

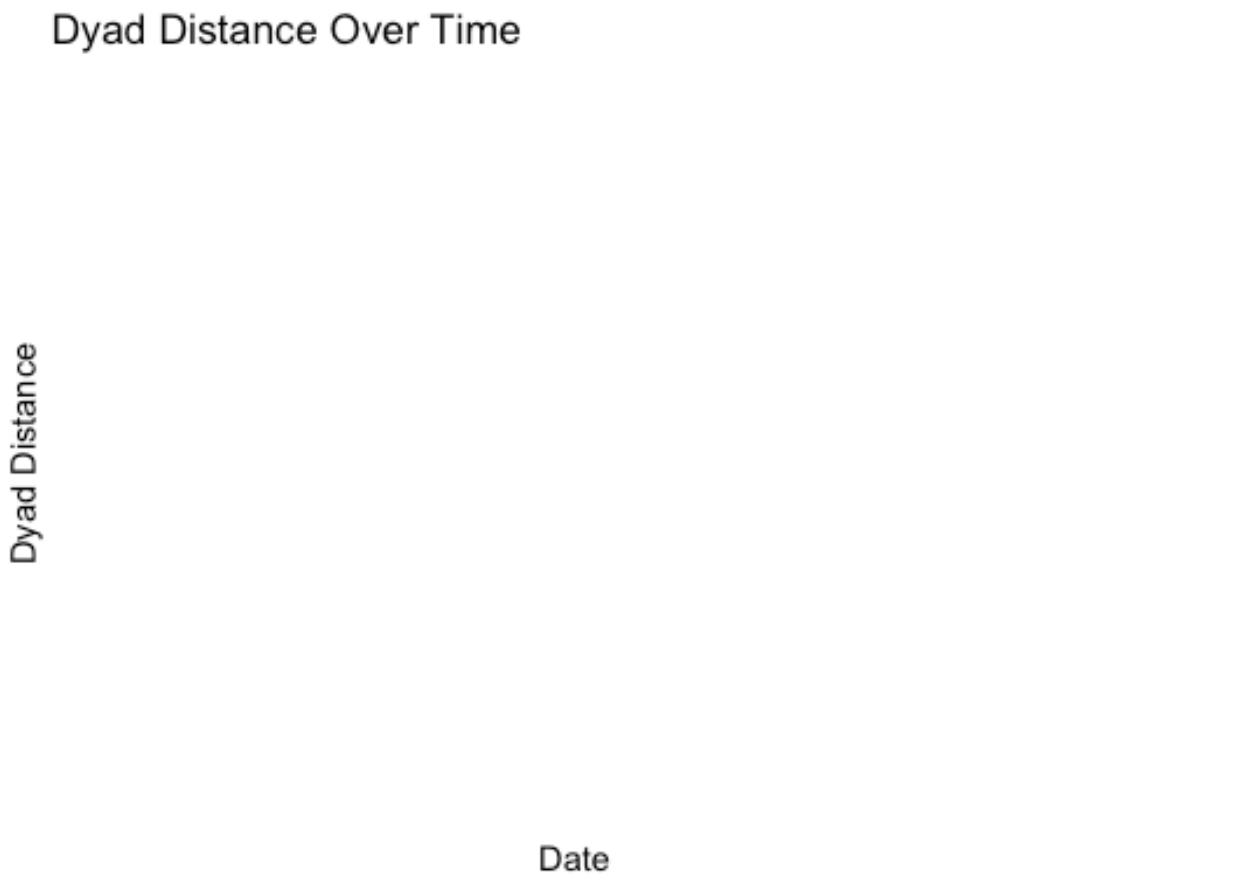
Exploratory Graph (To organise)

####Dyad, Distance & Date

- Trials of grpahs, I will have to check all of them
- My goal here is too see if each dyad have an general evolution of their dyad distance trough time and how many varaition do they have

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : at 19306

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : radius 0.04
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 19306

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 0.2

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 1616

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning: Computation failed in `stat_smooth()`
## Caused by error in `predLoess()`:
## ! NA/NaN/Inf dans un appel à une fonction externe (argument 5)
```




```
# Load required libraries
library(ggplot2)
library(dplyr)
```

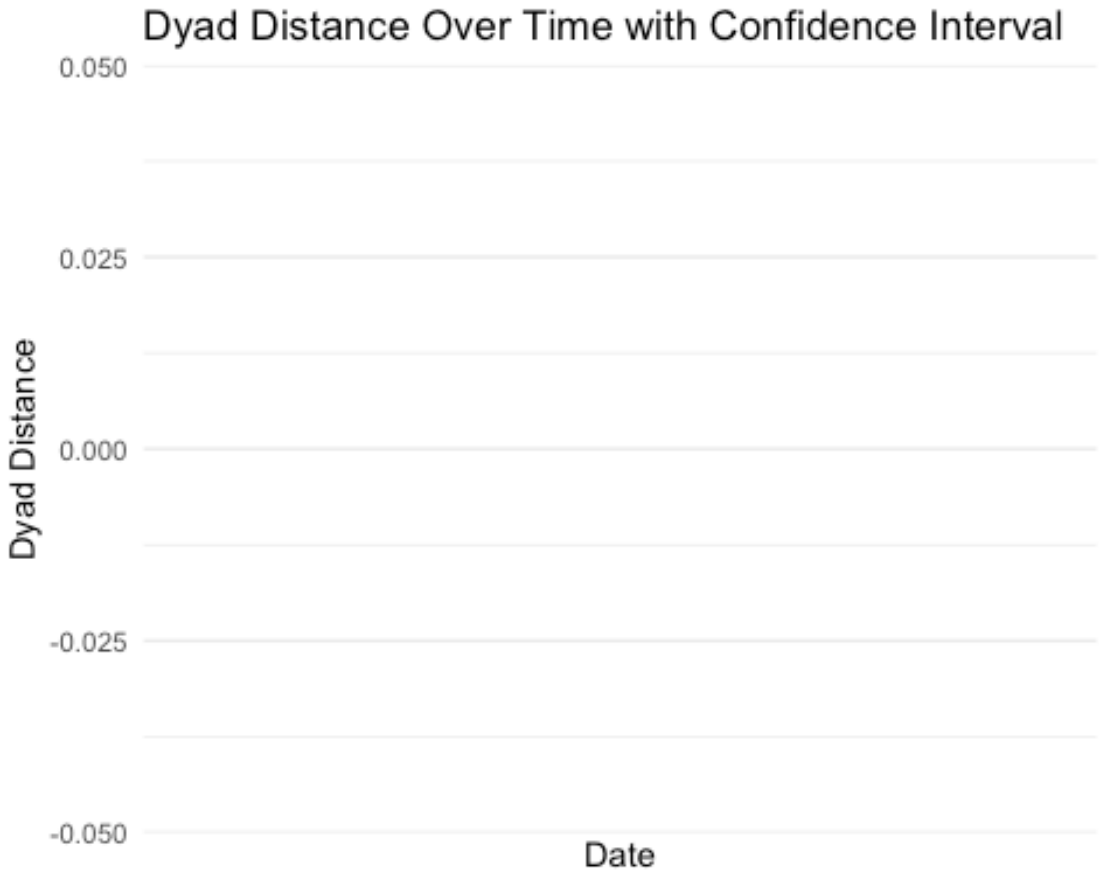
```r
# Ensure Date is in the correct format
Bex$Date <- as.Date(Bex$Date)

# Plot with smoothed lines and confidence intervals with improved readability
daily_plot <- ggplot(data = Bex, aes(x = Date, y = DyadDistance, group =
Dyad, color = Dyad)) +
  geom_smooth(method = "loess", se = TRUE, alpha = 0.2, size = 1) + #
Adjusted alpha and line size
  theme_minimal() +
  theme(legend.position = "right") + # Adjust legend position
  labs(x = "Date", y = "Dyad Distance", title = "Dyad Distance Over Time with
Confidence Interval", color = "Dyad") +
  expand_limits(y = c(0, NA)) # Expand y-axis limits if needed
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

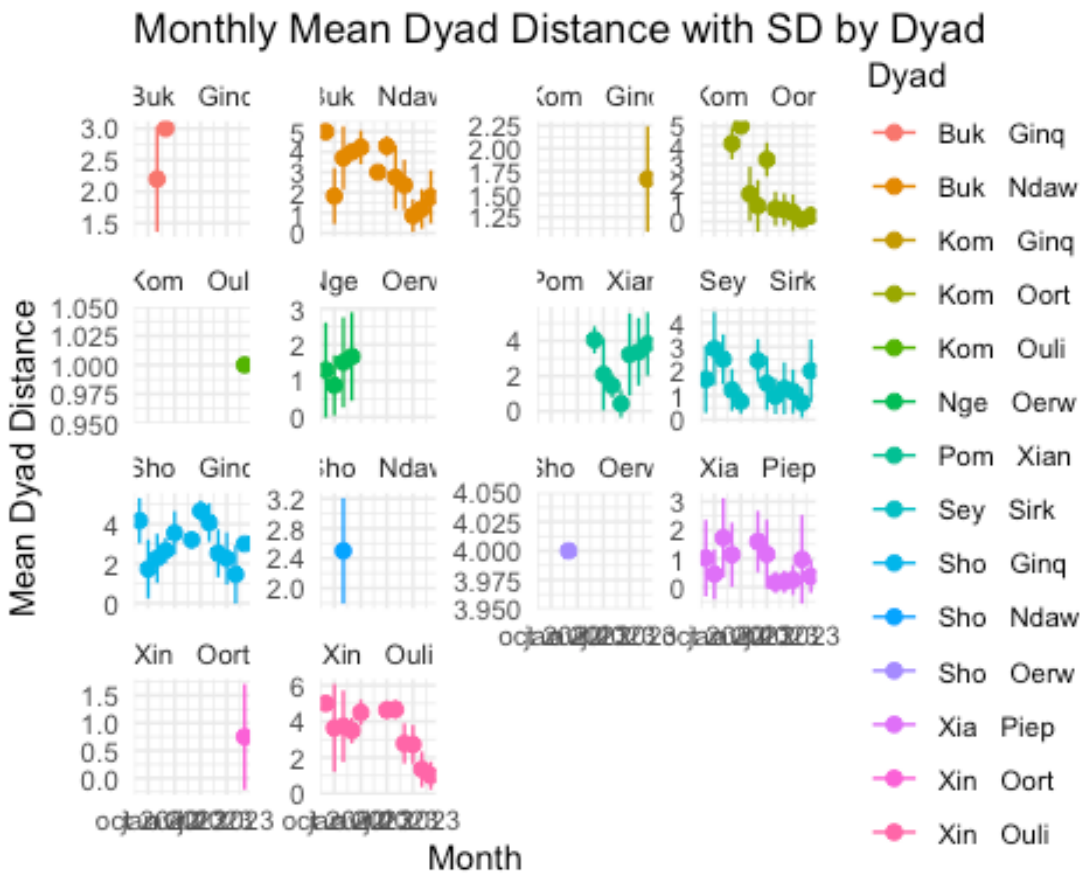
```r
# Print the daily plot
print(daily_plot)
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : at 19306

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : radius 0.04

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 19306

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 0.2

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 1616

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger
```

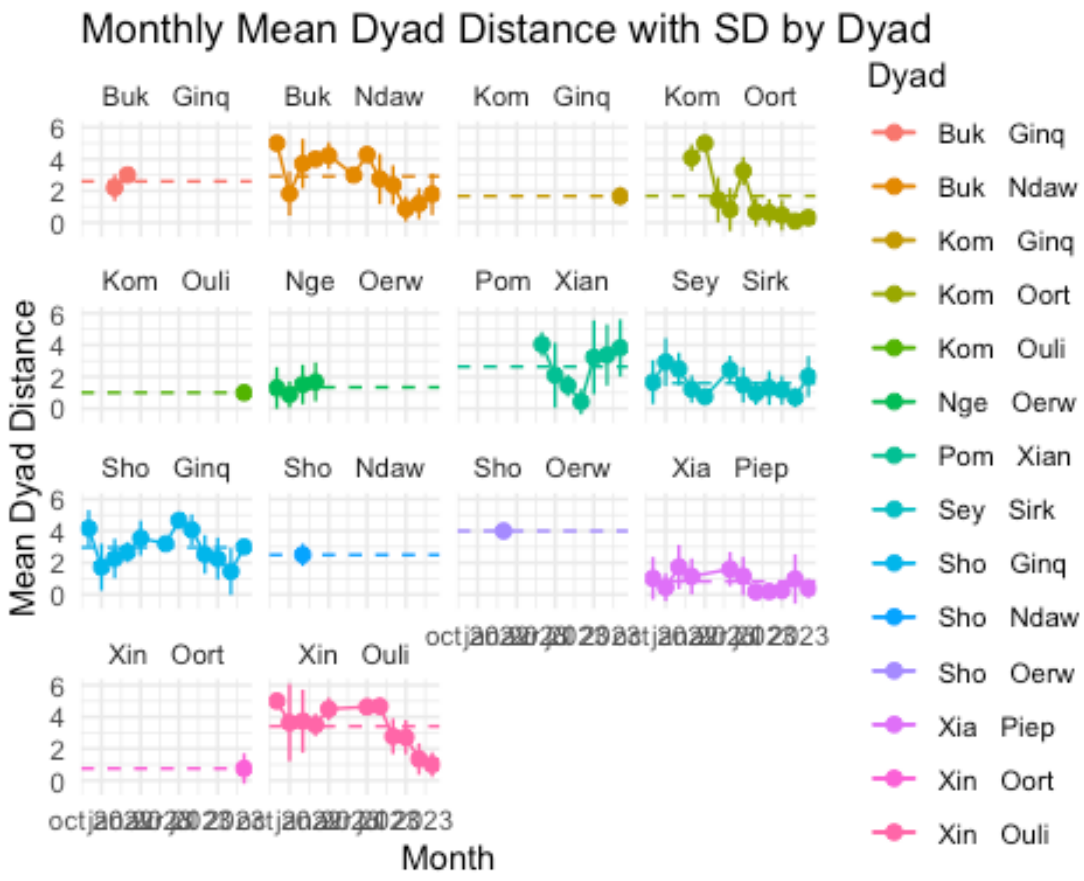
```
## Warning: Computation failed in `stat_smooth()`
## Caused by error in `predLoess()`:
## ! NA/NaN/Inf dans un appel à une fonction externe (argument 5)
```

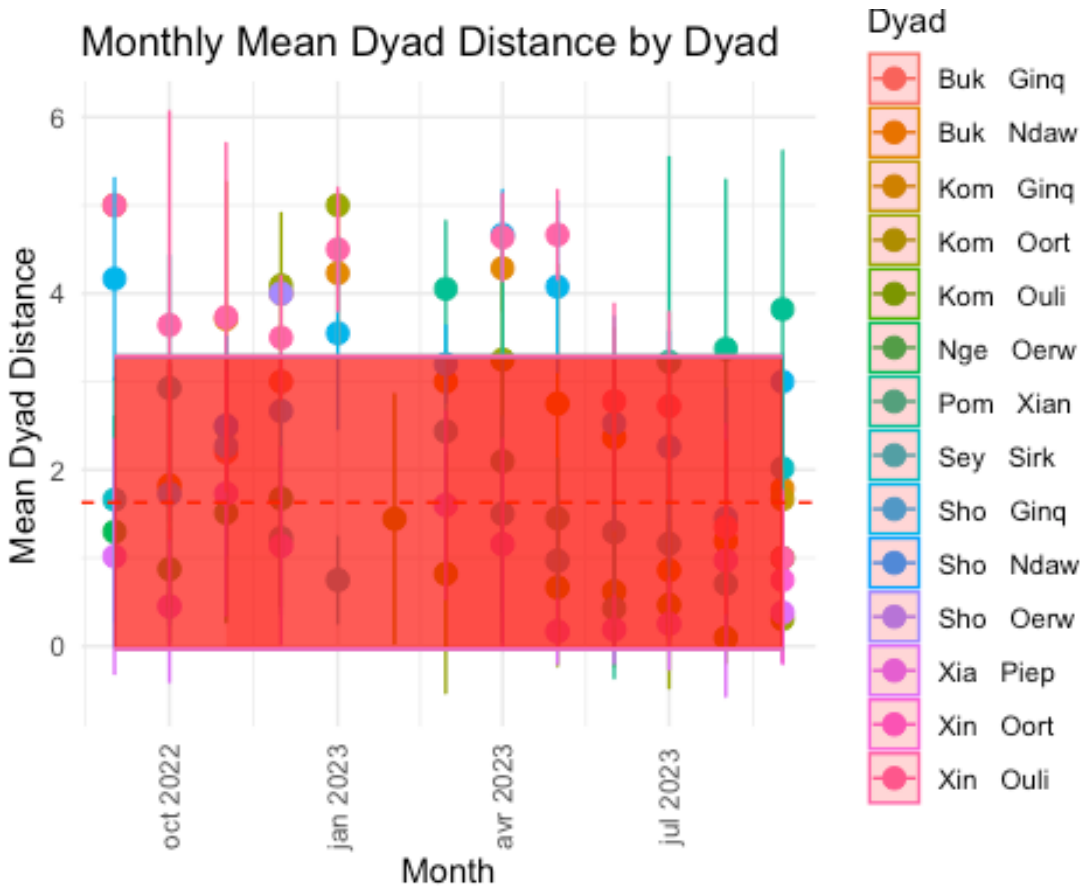



```
## `summarise()` has grouped output by 'Dyad'. You can override using the
## `.groups` argument.
```




```
## `geom_line()`: Each group consists of only one observation.
## ℹ Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## ℹ Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## ℹ Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## ℹ Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## ℹ Do you need to adjust the group aesthetic?
```


Monthly Mean Dyad Distance with SD by Dyad
Buk Ginq
Buk Ndaw
Kom Ginq
Kom Oort
Kom Ouli
Nge Oerw
Pom Xian
Sey Sirk
Sho Ginq
Sho Ndaw
Sho Oerw
Xia Piep
Xin Oort
Xin Ouli
Mean Dyad Distance
Month
Dyad
Buk Ginq
Buk Ndaw
Kom Ginq
Kom Oort
Kom Ouli
Nge Oerw
Pom Xian
Sey Sirk
Sho Ginq
Sho Ndaw
Sho Oerw
Xia Piep
Xin Oort
Xin Ouli



- I want to use the date to know **how many sessions** have been done with each dyads in my experiment. * I will create a variable called **Session** where **1 session = 1 day** * The data has values from the **14th of September 2022** until the **13th of September 2023** * I will also create a variable called **Trial** to know how many trials have been done with each dyad where **1 row = 1 trial**

```
* I may consider, in parallel of my hypothesis, to separate the data in *4
seasons* to make a preliminary check of a potential effect of seasonality.
Nevertheless the fact that we did not use anywithout      tools to mesure the
weather and the idea to make a categorization in 4 seasons without
considering the actua quite arbitrary. I may do it but with no intention to
include this in my scientific report.
l temperature, food quantitiy and other elements related to seasonailty make
this categorizationn a categorization where 12 months of          data will
be separated in 4 categories
```

- But before I may want to make a few changes already by merging **Male corn** and **Male placement corn** into ” Male corn” and maybe replacing all of the NA’s in “Other response” by response

#Lines to check unique values in MaleFemaleID to see if they are any problems with it # Unique values in MaleID unique_male_ids <- unique(Bex$MaleID)

Unique values in FemaleID

unique_female_ids <- unique(Bex$FemaleID)

Sections below are here for the organization of my paper and will be worked on once the data cleaning and exploration is done

3. Describing the data

4.Visualizing the data

5.Research question & Hypothesis

Research question

- What factors influence the rate at which individuals (vervets) learn to tolerate each other in a controlled box experiment?
- Ex: The rate at which individuals (vervets) learn to tolerate each other in a box experiment is influenced by social factors (audience, social network, behavior of the partner) and idioyncratic factors (age, rank)

Hypothesis

-
    1. Hypothesis about the Presence of High-Ranking Individuals:

The presence of a higher number of high-ranking individuals in the audience will negatively correlate with the level of tolerance achieved among vervets in the box experiment. This is expected to result in higher frequencies of aggressive behaviors, intrusions, and loss of interest, particularly from lower-ranking individuals.

-
    2. Hypothesis about Partner Agonistic Behaviors:

Vervets tolerance levels in the box experiment will be influenced by their partner's display of agonistic behaviors. Specifically, partners who exhibit more frequent agonistic behaviors towards their partner will lead to decrease in their motivation to participate in future trials.

-
    3. Hypothesis about the Establishment of an Optimal Distance:

During the box experiment, vervet dyads will establish an "optimal" distance for interaction, characterized by a higher frequency of tolerance compared to other distances. This optimal distance is expected to signify that the individuals tolerate each other more effectively at this specific proximity .

-
    4. Hypothesis about Age and Rank:

The age and rank of individual vervets within the group will influence the success of the trials in the box experiment. Specifically, older and higher-ranking individuals are expected to exhibit lower rates of success compared to dyads consisting of younger and lower-ranked individuals. This decrease in success is anticipated to be associated with a higher frequency of aggressive behaviors displayed by older and higher-ranking individuals towards their partners. (I'm not sure this hypothesis makes sens, I have the feeling age and rank must have an influence but I don't know how to put it, I will think about it)

- 5. Hyptohesis about seasonality

Seasonality is expected to impact the motivation of vervet dyads to participate in the box experiment. We hypothesize that dyads will have lower motivation, as indicated by a reduced number of trials, during the summer months compared to the winter months. This difference in motivation is likely influenced by temperature and food availability. To test this hypothesis, we will categorize the data into four seasonal periods, each spanning four months, and analyze whether there is a significant effect of seasonality on the motivation to engage in the trials.

6.Statistical tests and analisis of the data

Statistical tests

- **Hypothesis 1**: Influence of High-Ranking Individuals

Variables Needed:

DyadResponse (specifically, "aggression" responses) **Amountaudience** (to measure the number of individuals in the audience) **Audience…15** (to identify the names of individuals in the audience for calculating dominance ranks) **Elo rating** of the individuals based on the ab libitum data collected in IVP (which I have to calculate asap)

Statistical Analysis:**Logistic Regression**, as it could analyze the influence of high-ranking individuals on the occurrence of aggression in dyad responses. This will help determine whether the presence of high-ranking individuals affects the likelihood of aggression.

- **Hypothesis 2**: Impact of Partner's Agonistic Behaviors

Variables Needed:

- **DyadResponse** (specifically, "aggression" responses)
- **MaleagressF** (male's aggression towards female)
- **FemaleaggressM** (female's aggression towards male)

Statistical Analysis: **Logistic Regression** as it could be used to assess how the occurrence of aggression in dyad responses is influenced by the partner's gender-specific agonistic behaviors.

- **Hypothesis 3**: Identification of an Optimal Interaction Distance

Variables Needed:

- **DyadDistance** (distance between boxes)
- **Tolerance** (as a binary outcome)

Statistical Analysis: **generalized Linear Model (GLM)** to investigate whether there is an optimal distance that leads to a higher likelihood of tolerance (Tolerance = 1).

- **Hypothesis 4**: Role of Age and Rank

Variables Needed:

- **Tolerance** (as a binary outcome)
- **Male and Female** (to identify individuals' ages and ranks)
- **Dyad** (to link individuals to dyads)
- **Birthdate** to calculate the age of each individual

Statistical Analysis: **Logistic Regression** Logistic regression can be employed to determine whether the age and rank of individual vervets within dyads have an impact on the likelihood of tolerance (Tolerance = 1).

- **Hypothesis 5**: Influence of Seasonality

Variables Needed:

- **Date** (to categorize data into seasons)
- **Trial** (to count the number of trials in each season) and the data for at least 365 days so i can separate the data in 4 (1 year = 4 seasons = 12*4 month) to see if they may be an effect of seasonality on the motivation (amount of trials) of the dyads

Statistical Analysis:

ANOVA or Kruskal-Wallis Test: Depending on the distribution of your trial data, you can use either ANOVA (if the data are normally distributed) or the Kruskal-Wallis test (for non-normally distributed data) to assess the impact of seasonality on the number of trials. If significant differences are found, you can follow up with post-hoc tests to identify which seasons differ from each other. Please note that the effectiveness of these analyses may depend on the distribution of your data and specific research objectives. You may also consider conducting exploratory data analysis (e.g., visualization) to gain a better understanding of your dataset before performing these analyses. Additionally, if you have specific questions about data preprocessing or variable transformations, feel free to ask for further guidance. –> I took this from ChatGPT, I have to look more into it

REMARKS: So here are a few updates I made in the document. I also planned to send my cleaned data to Radu (the statistician of UNINE) as he was keen to help me find the right test. Of course I will also look again in Bshary's and Charlotte's work with the boxes and improve these suggestions that are quite simple for now

Also I still have to clean the last grpahs about male/female aggression as I didn't finish that yet. I juste wanted to share my hypothesis and ideas for statistics so I can soon go into the "serious" work

Anyway, thank you in advance for your help <3

Michael

7. Plotting the results of the analysis

9. Interpretation of the results

10. Comeback on the research question and hypothesis

11. Bibliography

12. Organization for my paper

- Introduction
  - Tolerance humans, primates
  - Apes vs monkeys / Captivity vs Wild
  - IVP: Wild habituated vervets, experiments possible
  - Paper Bshary, Canteloup… Prolongation study
  - Relevance idea/topic research
  - Research question & hypothesis

But: intro need triangle shape: broad to narrow end wiht research question> tolerance importance > animal reign, actual knowledge/ direction knowledge we need > show how my experiment goes in that way How to adress the gap, answer with research question

Then explain why choosing vervet monkeys, (IVP in methods), sociality, experiments made

- Methods
  - IVP, research area, (goal, house, type people)
  - Population: groups, dyads, male/female, ranks..
  - Box material: boxes, remotes, batteries, camera, tripod, corn (no marmelade ;), (water spray, security reason, non agressive way to select individuals and not engage with mokeys when reachrging boxes with corn), pattern, previous distances, tablets, box experiment form
  - Tablets
  - (No observers mentionned)
  - Habituation boxes > individuals trained to recognice boxes, they have differernt levels of habituation

- Patterns > appendix, mention similar to habituation, use to recognize box but efficieny depeds of experience)
- Selection dyads > assigment from elo rating (different rank), if above average bond no dyad made, if not possible, availibilty of monkey also factor !! Non random can be a problem, think about why and how you selected data We created variations in dyads made by different sex, rank and not above average bonde (calculate bondeness)
- Amount corn, do you want to mention it> maybe important Calculate corn during and placement cf paper on corn /food motivation
- Corn (daily intake vervet % made from corn, cf site we saw, cf screenshot, comapre paper previousely made an all)
- 1st dyad trial (BD) > appendix
- Videos > details appendix
- Finding dyads > appendix
- Placement to attract them > meniton if statiscial made on placement corn
- Trials (1 session = max 15 trials/in total) (session could be broken in different sub sessions to reach 15 trials max)
- If agression > 1m / If 2x tolerance < 1m , also if not approaching > 1m (if no tolerance increase distance except if intrusion) (borgeaud > expectation fo aggression)
- Time of the day > appendix
- Territory? > appendix
- Amount sessions p day/week, how we chose the moment to follow them >appendix
- Problems/ unplanned events: weather, BGE's, not finding the monkeys (group, dyad or individual), dispersal of males, river crossing, inacessibility (experiments or boxes), low vision (experiments or monkeys),> appendix
- (Where do i mention the confounding variables?) > look in litterature, if something that could affect and already reported in papers check, oterhwise exclude "normal life" factors for both monekys and Experimenter
- Types of experimental plan
- Statistical tests (for each hypothesis)

- Analysis

- Results
- Interpretation
- Conclusion

Glossary

- **Tolerance**: Tolerance: An individual has an encounter with a conspecific and can freely leave but remains in the encounter without acting aggressively toward the conspecific. (Pisor & Surbeck, 2019)
- **Agression**
- **Session**
- **Trial**
- **Group**: In the Primate order, groups are individuals "which remain [physically] together in or separate from a larger unit" and interact with each other more than with other individuals.6 This definition does not cover all uses of the word "group" in the social sciences (e.g., human identity groups who identify with a common name or symbol may or may not interact with one another more frequently than with other individuals). Because of this ambiguity, we use the word "community" when referring to humans to better capture the notion of spatial proximity, per Ref. 54. Members of the same group are referred to as "same-group" and those from another group "extra-group." (Pisor & Surbeck, 2019)

Bibliography

• Pisor, A. C., & Surbeck, M. (2019). The evolution of intergroup tolerance in nonhuman primates and humans. Evolutionary Anthropology: Issues and ReViews. Advance online publication. https://doi.org/10.1002/evan.21793 (Pisor & Surbeck, 2019)

Annex

Annex 1 : View of the dataset when imported - First 6 entries of each variable

- We can see here the brief View of the **original dataset** names **BoxEx**when i initially imported it as seen in **section 0: Opening data**

First Few Entries (continued below)

Date	Time	Data	Group
2022-09-27	1899-12-31 09:47:50	Box Experiment	Baie Dankie
2022-09-27	1899-12-31 09:50:07	Box Experiment	Baie Dankie
2022-09-27	1899-12-31 09:53:11	Box Experiment	Baie Dankie

Date	Time	Data	Group
2022-09-27	1899-12-31 09:54:28	Box Experiment	Baie Dankie
2022-09-27	1899-12-31 09:55:19	Box Experiment	Baie Dankie
2022-09-27	1899-12-31 09:56:56	Box Experiment	Baie Dankie

Table continues below

GPSS	GPSE	MaleID	FemaleID
-28.010549999999999	31.191050000000001	Nge	Oerw
-28.010549999999999	31.191050000000001	Nge	Oerw
-28.010549999999999	31.191050000000001	Nge	Oerw
-28.010549999999999	31.191050000000001	Nge	Oerw
-28.010549999999999	31.191050000000001	Nge	Oerw
-28.010549999999999	31.191050000000001	Nge	Oerw

Table continues below

Male placement corn	MaleCorn	FemaleCorn	DyadDistance	DyadResponse
NA	3	NA	2m	Tolerance
NA	3	NA	2m	Tolerance
NA	3	NA	1m	Tolerance
NA	3	NA	1m	Tolerance
NA	3	NA	0m	Tolerance
NA	3	NA	0m	Tolerance

Table continues below

OtherResponse	Audience	IDIndividual1	IntruderID
NA	Obse; Oup; Sirk	NA	NA
NA	Obse; Oup; Sirk	NA	NA
NA	Oup; Sirk	NA	NA
NA	Sirk	NA	NA
NA	Sey; Sirk	NA	NA
NA	Sey; Sirk	NA	NA

Table continues below

Remarks

Remarks
NA
NA
Nge box did not open because of the battery. Oerw vocalized to MA when he ap to the box to open it.
Sey came to the boxes once they were open
NA
NA

Observers	DeviceId
Josefien; Michael; Ona; Zonke	{7A4E6639-7387-7648-88EC-7FD27A0F258A}
Josefien; Michael; Ona; Zonke	{7A4E6639-7387-7648-88EC-7FD27A0F258A}
Josefien; Michael; Ona; Zonke	{7A4E6639-7387-7648-88EC-7FD27A0F258A}
Josefien; Michael; Ona; Zonke	{7A4E6639-7387-7648-88EC-7FD27A0F258A}
Josefien; Michael; Ona; Zonke	{7A4E6639-7387-7648-88EC-7FD27A0F258A}
Josefien; Michael; Ona; Zonke	{7A4E6639-7387-7648-88EC-7FD27A0F258A}

```beforeNA
```