BEX 223 MAungKyaw

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 nd the **pander** package for better presentation
- The **dplyr** package was installed for better manipulation of the data as filtering or creating new variables
- Then, I installed the **readxl package** to import my dataset which is called **Box Experiments.xls**

1.Explore the data

Description of the initial datset - "Boxex"

```
## Glimpse of the 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 **glimpse** function to display a summary of my dataset
- I have **20 columns** (here variables) and **2795 rows** (here trials)
- I will now make a brief summary of each variables and their use before creating a new dataframe (df) with my variables of interest
- The highlighted variables are the ones I will use for my new df. I will then **clean the data** before heading to the **statistical analysis**
- The variables we have in this dataset are the following:

Variables of Boxex

- **Date** : "Date"cis in ac**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, 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 measure 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
- **Time** : "Time is coded" in a **POSIXct** format
  - I do not plan to use this variable but we can see that "Time" has 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, I would need to correct the incorrect display of the date in the dataset.
  - This variable could also be usefull to seen 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**
  - I will describe the variable again after cleaning it
- Data : chr "Data" is coded as **character**
  - It describes **what type of data** has been entered the software **cybertracker**. We installed the software on tablets to record the different behaviours of vervet monkeys in our research center
  - In our case, the whole dataset is coded as **Box Experiment** as it was the type of behaviour that we were recording
  - For this reason we can remove this column as the information is unecessary
- 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 theamount 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 gives 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 "FealeID" is coded as **character**
  - It gives 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 monkey of the dyad before the trials**
  - Within a session it happened that we gave more placement corn t to attract the monkeys again to the boxes. This lead to an update of the number in the same session. The number that 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
  - I will change the format of the varaible to numerical
- **MaleCorn** : dbl "MaleCorn" is coded in r as **double**
  - It gives the same information as in ***male placement corn***
  - I will add the data from "male placement corn" into this one
  - I will change the format of the varaible to numerical
- **FemaleCorn** : dbl The data is coded in r as **double**
  - It gives the **amount of corn given to the monkey before the trials**
  - It works in the same way as "male placement corn"/"MaleCORN"
  - I will change the format of the varaible to numerical
- **DyadDistance** : chr The data is coded in r as **character**
  - I gives the **distance for each trial** 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 insterad of character
- **DyadResponse** : chr The data is coded in r as **character**
  - It gives the **response for each trial** we have done with the dyads
  - The different behaviours were: **Distracted**.**Female aggress male**, **Male aggress female**, **Intrusion**, **Loosing interest**, **Not approaching**, **Tolerance** and **Other**
  - I will create diatomic variable for each behaviour in **DyadResponse** to have the frequency of each behaviour
  - As multiple behaviours could be found in a single cell I will create a hierarchy to reduce the amount of behaviours assigned to each trial (if there is more than one)
  - Projection of the hierarchy (to be modified)
    - 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 **another behaviour from the one found in Dyad Response** (so if it is not tolerance, aggression, intrusion, loosing interest, not approaching, distracted)
    - I will have to look at every **OtherResponse** and rename them
    - 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
- **Audience** : chr "Audience" is 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, 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, show aggression or lost interest** during a trial
    - I will have to look at it to see how often these behaviours occured and from which individual
    - I will consider how to use this vaiable in more detail later
- **IntruderID** : chr "IndtruderID" is coded as **character**
    - It gives the **name of the individual that did an intrusion during a trial**
- **Remarks** : chr The data is coded in r as **character**
    - It gives supplementary information concerning the experiment when unusual behaviors occurred or when we considered to add informations on the trial
- 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
- 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. Cleaning the data

Select variables

- Since I do not want to work with this whole dataset, I'm gonna select the variables of interest using "select"

- 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
- I will then keep Date, MaleID, FemaleID, Male placement corn, MaleCorn, FemaleCorn, DyadDistance, DyadResponse,OtherResponse, Audience. . . 15, IDIndividual1, IntruderID, Remarks

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**

Table 1: 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
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 2: 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 3: 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 4: Table continues below

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

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

Table 5: Table continues below

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}