THE AMOUNT OF DETAILS IN TRUTHS AND LIES ABOUT PAST AND FUTURE WEEKEND ACTIVITIES

Sally A.M. Hogenboom - 11377909

16-12-2017

Available Data

- File 1 - LieDetection_Data.csv
  - Contains demographics, veracity judgement, exclusion criteria, transcriptions, confidence of deceit, frequency of lie.
- File 2 - LieDetection_NER_processed.csv
  - Transcriptions in long format -> allows for Named Entity Recognition (NER;https://spacy.io)
  - Unique and Non-Unique Frequency counts of the NERs...
    - Person
    - Location
    - Time
    - Facility
    - GPE
    - Product
    - Date
    - Money

Meaning of Column Headers

1. `Participant.Nr.` > unique participant number, corresponds with filenames of voice tracks.
2. `Condition.Nr.` > Did participants talk about future or past weekend activities?
3. `Location` > Where the participant Data_All was collected
4. `Date.Time` > When the Data_All was collected
5. `Age` > Age of the participants during Data_Allcollection
6. `Gender` > Self-identified gender
7. `Student..Y.N` > Are the participants students?
8. `English.Proficiency` > On a level from 1 (not at all) to 5 (excellent) how do you rate your own level of English proficiency?
9. `Ground.Truth.1` > 1 Sentence Description of Past/Future Saturday activities
10. `Ground.Truth.2` > 1 Sentence Description of Past/Future Sunday activities
11. `OrderOfLie` > Did participants start by lying or telling the truth? Included as covariate, to detect any effect of modelling after the first description
12. `Veracity.Accuracy` > Was the 3rd researcher succesful in destinguishing the truth from the lie?
13. `Veracity.Coding` > Numerical values of 12
14. `Judgement.By` > Who provided the veracity judgement?
15. `Frequency.of.Lie` > Self-rated; on a level from 1 (never) to 10 (all the time), how often do the activity that you lied about (e.g., how often do you go to the gym?).
16. `Confidence.of.Deception` > On a scale from 1(not at all) to 10 (definitely) how sure are you that you have deceived the judge? In other words, higher scores indicate that the participants believe their statements to be highly similar

17. `Saturday` > Transcriptions of the statement provided about the saturday
18. `Sunday` > Transcriptions of the statement provided about the sunday
19. `NOTES` > Any notable events that occurred during Data_All collection or basis for the veracity judgements.
20. `Transcriptions` > Any notes regarding the transcriptions
21. Named Entities (Sat_ = Saturday & Sun_ = Sunday)
    a. `_Veracity` > Whether the description is a truth or a lie
    b. `_Person_NU` > Non-unique person entities (people)
    c. `_Location_NU` > Non-unique location entities (where)
    d. `_Time_NU` > Non-unique time entities (smaller than a day)
    e. `_GPE_NU` > Geographic locations
    f. `_Product_NU` > Non-unique product entities (e.g., Nintendo)
    g. `_Date_NU` > Non-unique date entities (e.g., monday)
    h. `_Money_NU` > Non-unique money entities (e.g., 500 dollars)
22. `Total_` > Total number of named entities in truthful or deceptive statements
23. `Predictive_Accuracy` > Veracity judgement made based on criteria that lies contain more details than truthful statements

Predictive Accuracy

Confirmatory analysis indicated that lies, on average, contain more Named Entities than truthful statements. If we adopt this criteria as part of exploratory predictive modelling we find the following frequencies.

```
##              Future Past Total
## Correct          15   18    33
## Inconclusive     16   11    27
## Mistaken          8    9    17
## Total            39   38    77
```

The percentage accuracy attained from conclusive judgements (N = 50) is:

```
## [1] "66 %"
```

Exclusion Criteria

Participants were excluded based on:

- Inadherence to the instructions (N = 2)
  - PP. 5
  - PP. 21
- Change of protocol (N = 15)
  - PP. 63 - PP. 77

```
Data_Excluded <- Data_All[-(c(5,21,63:77)),]
```

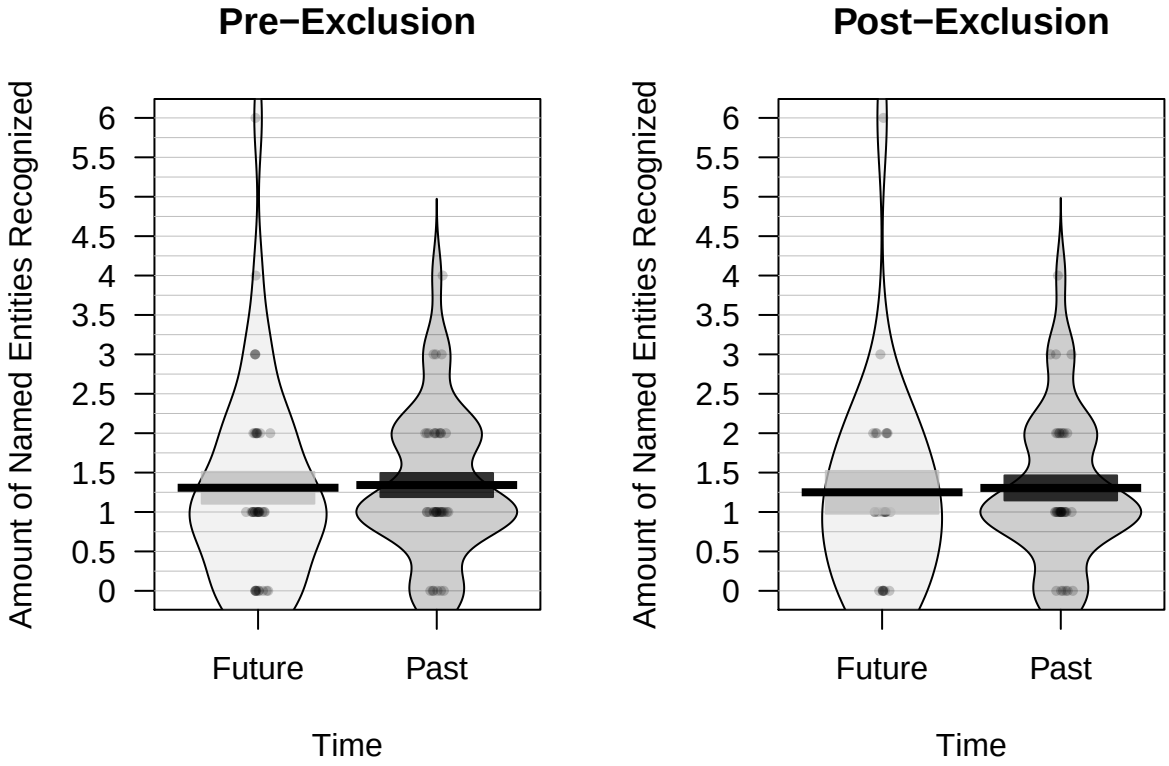
Predictive Accuracy After Exclusion

```
##              Future Past Total
## Correct          11   18    29
## Inconclusive      9   10    19
## Mistaken          4    8    12
## Total            24   36    60
```

```
## [1] "71 %"
```

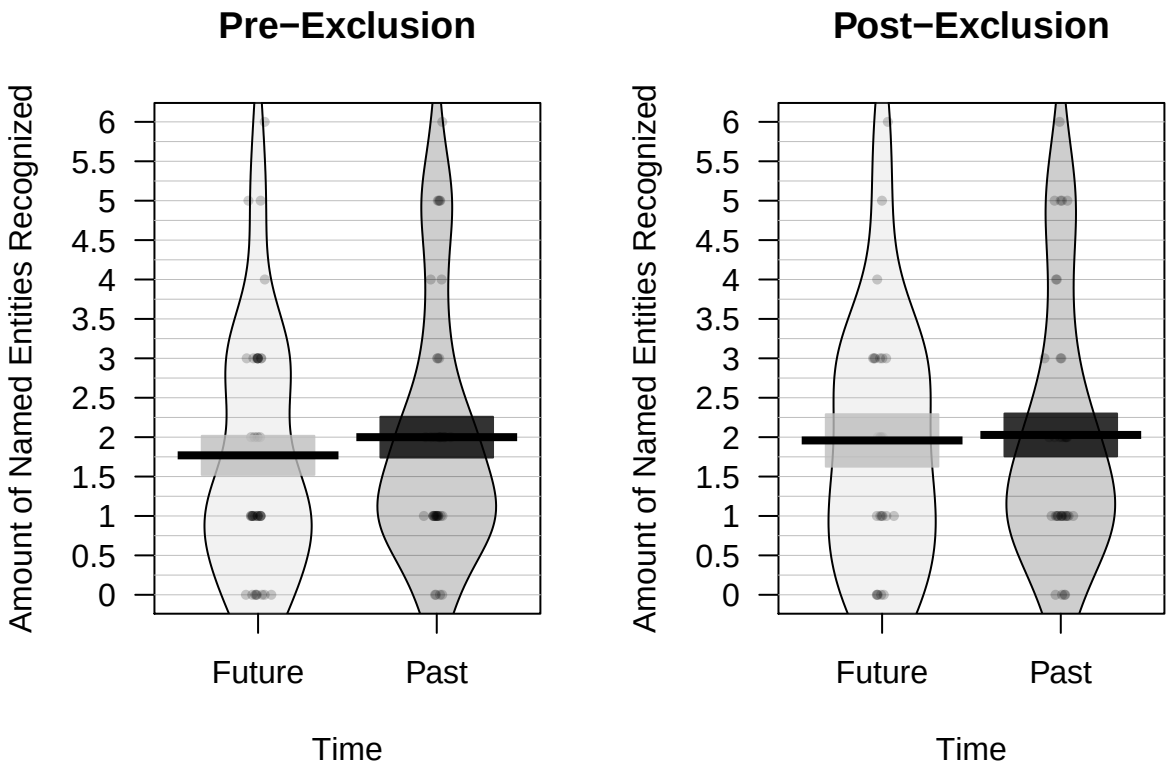
Visualization of Data

Truthful Statements



The amount of recognized named entities for truthful statements about past and future weekend activities. On the left the data is presented **before** exclusion, and on the right **after** exclusion of data. The black line represents the group mean, the box the standard error of the mean, and the spread of the data is outlined (Kampstra, 2008)

Deceptive Statements



The amount of recognized named entities for deceptive statements about past and future weekend activities. On the left the data is presented **before** exclusion, and on the right **after** exclusion of data. The black line represents the group mean, the box the standard error of the mean, and the spread of the data is outlined (Kampstra, 2008)

Processed Data

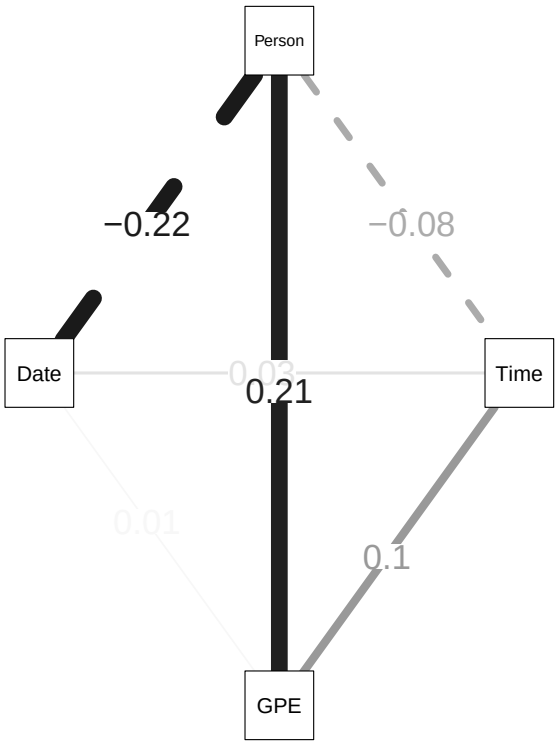
The datafiles including newly created columns (e.g., total amount of NEs in truthful descriptions) are available as:

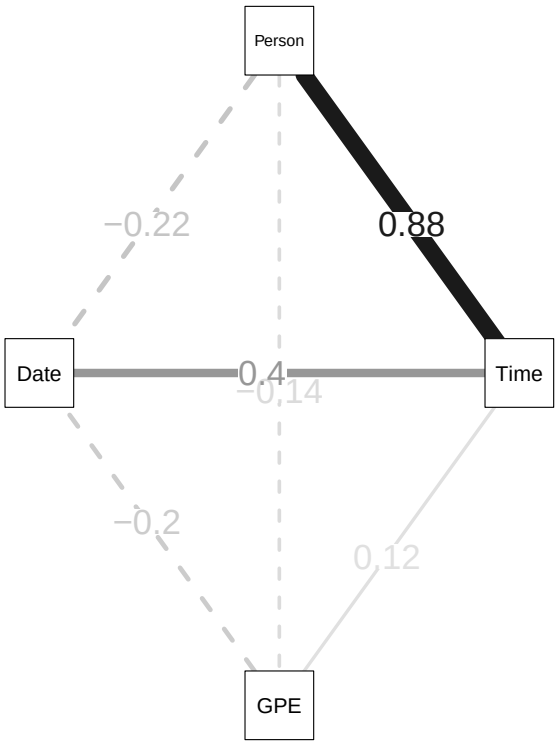
- All collected data - LieDetection_CombinedData_All.csv
- Data after excluding of participants - LieDetection_Combined_Excluded.csv

Network Analysis

Truthful Statements

```
## Loading required package: psych
## Loading required package: qgraph
```





Deceptive Statements