Data Cleaning

Kennedy Zapalac

2023-10-03

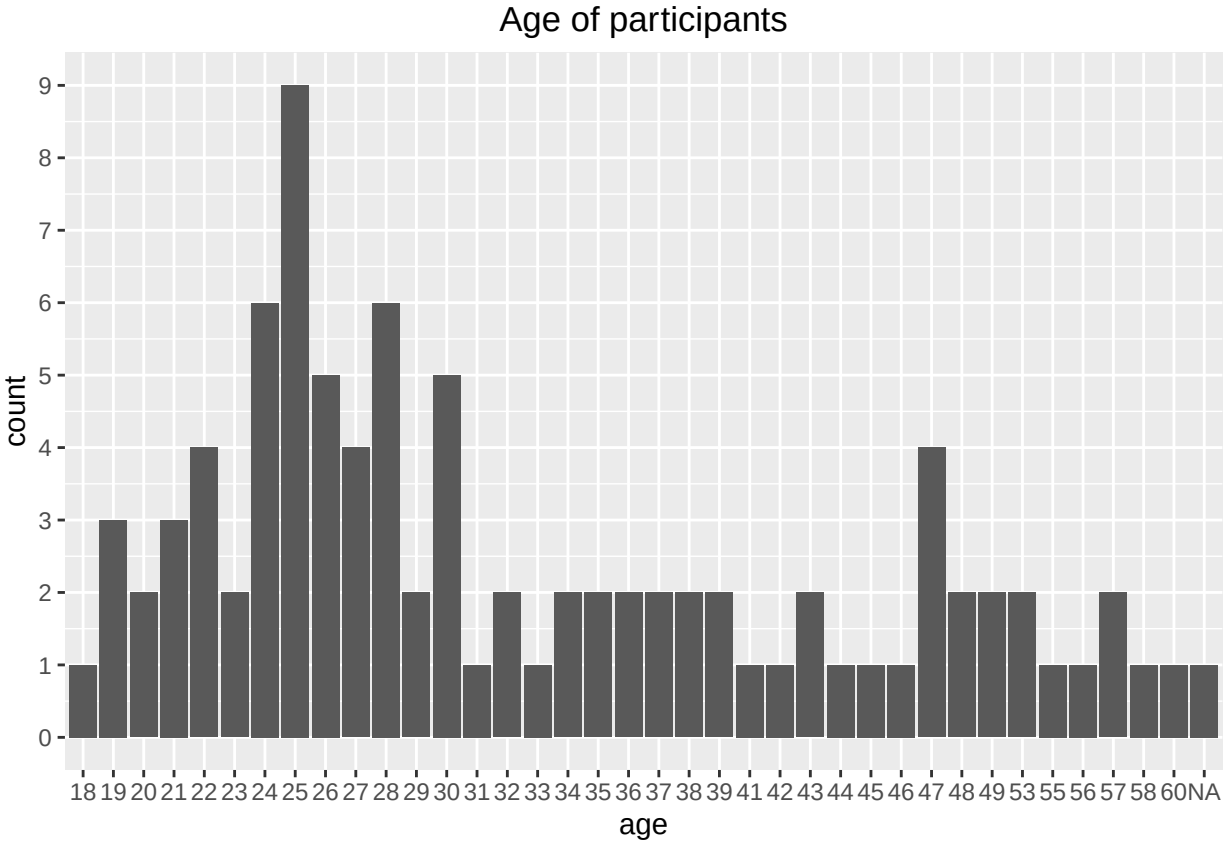
After cleaning the data, here is what it looks like:

```
## Rows: 216
## Columns: 44
## $ startDate             <chr> "5/22/13 16:54", "5/23/13 16:15", "5/23/13 16:21~
## $ endDate               <chr> "5/22/13 17:08", "5/23/13 16:20", "5/23/13 16:30~
## $ finished              <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1"~
## $ consent               <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1"~
## $ age                   <chr> "19", "26", "28", "38", "18", "21", "20", "23", ~
## $ gender                <fct> female, female, male, male, male, male, female, ~
## $ self_disc             <chr> "I'm a generally happy person and I like to make~
## $ priming_response      <chr> "We had a date night for my sorority a couple we~
## $ att_check             <chr> "She and her boyfriend broke up", "The breakup w~
## $ note                  <chr> "My boyfriend and I broke up a little over a mon~
## $ pt_1                  <dbl> 1, 2, 2, 2, 3, 4, 1, 1, 1, 1, 2, 1, 1, 2, 5, 1, ~
## $ pt_2                  <dbl> 7, 6, 6, 6, 5, 4, 7, 7, 6, 6, 6, 7, 7, 5, 6, 7, ~
## $ pt_3                  <dbl> 2, 1, 2, 1, 4, 4, 1, 1, 1, 1, 3, 1, 1, 4, 3, 1, ~
## $ pt_4                  <dbl> 7, 5, 6, 7, 5, 4, 7, 6, 7, 7, 4, 7, 6, 3, 6, 5, ~
## $ pt_5                  <dbl> 1, 2, 2, 1, 2, 2, 1, 1, 1, 1, 6, 1, 1, 1, 5, 1, ~
## $ emp                   <chr> "6", "5", "5", "6", "3", "4", "7", "7", "7", "7"~
## $ dg                    <dbl> 0, 5, 0, 20, 0, 5, 2, 12, 0, 25, 0, 0, 10, 5, 0,~
## $ lp_scale_1_1          <dbl> 7, 6, 4, 6, 6, 4, 7, 6, 6, 6, 3, 6, 5, 4, 5, 5, ~
## $ lp_scale_1_2          <dbl> 4, 7, 4, 1, 5, 4, 3, 6, 5, 7, 3, 7, 5, 4, 3, 7, ~
## $ lp_scale_1_3          <dbl> 5, 6, 6, 5, 5, 4, 6, 6, 7, 6, 5, 5, 6, 6, 5, 5, ~
## $ lp_scale_1_4          <dbl> 6, 5, 6, 7, 4, 4, 7, 6, 6, 7, 6, 6, 6, 5, 5, 6, ~
## $ lp_scale_1_5          <dbl> 5, 5, 6, 7, 5, 4, 7, 6, 7, 7, 6, 4, 5, 5, 5, 3, ~
## $ lp_scale_1_6          <dbl> 7, 7, 4, 6, 3, 4, 3, 7, 5, 7, 6, 5, 5, 4, 6, 7, ~
## $ lp_scale_1_7          <dbl> 4, 4, 6, 6, 6, 4, 3, 5, 6, 7, 3, 5, 5, 2, 2, 6, ~
## $ lp_scale_1_8          <dbl> 7, 5, 6, 6, 7, 4, 7, 6, 7, 6, 6, 6, 6, 6, 5, 6, ~
## $ lp_scale_1_9          <dbl> 3, 5, 6, 7, 4, 4, 6, 6, 7, 6, 6, 5, 5, 5, 5, 2, ~
## $ lp_scale_2_1          <dbl> 6, 6, 6, 7, 7, 4, 7, 6, 6, 6, 5, 7, 7, 4, 5, 6, ~
## $ lp_scale_2_2          <dbl> 6, 7, 6, 7, 6, 4, 7, 6, 7, 7, 2, 7, 6, 4, 6, 5, ~
## $ lp_scale_2_3          <dbl> 4, 5, 6, 7, 6, 4, 4, 2, 6, 6, 3, 6, 5, 3, 3, 6, ~
## $ lp_scale_2_4          <dbl> 5, 6, 6, 7, 5, 4, 5, 6, 6, 6, 4, 7, 5, 5, 5, 6, ~
## $ lp_scale_2_5          <dbl> 7, 7, 6, 6, 7, 4, 7, 6, 7, 6, 5, 6, 6, 5, 5, 6, ~
## $ att_check2            <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ lp_scale_2_7          <dbl> 3, 6, 6, 7, 6, 4, 6, 2, 5, 5, 5, 7, 6, 4, 5, 7, ~
## $ lp_scale_2_8          <dbl> NA, 5, 4, 6, 5, 7, 6, 6, 7, 5, 7, 6, 6, 5, 5, 6,~
## $ includeOrExclude      <fct> NA, include, include, include, include, exclude,~
## $ priming_received      <chr> "performance", "learning", "performance", "learn~
## $ inOrOut               <chr> "in", "out", "out", "out", "in", "out", "in", "i~
## $ condition             <chr> "perf_in", "learn_out", "learn_out", "learn_out"~
## $ pt_1_reversed         <dbl> 7, 6, 6, 6, 5, 4, 7, 7, 7, 7, 6, 7, 7, 6, 3, 7, ~
```

```
## $ pt_3_reversed         <dbl> 6, 7, 6, 7, 4, 4, 7, 7, 7, 7, 5, 7, 7, 4, 5, 7, ~
## $ pt_5_reversed         <dbl> 7, 6, 6, 7, 6, 6, 7, 7, 7, 7, 2, 7, 7, 7, 3, 7, ~
## $ PT_avg                <dbl> 6.8, 6.0, 6.0, 6.6, 5.0, 4.4, 7.0, 6.8, 6.8, 6.8~
## $ trait_learning_avg    <dbl> 5.571429, 5.625000, 5.750000, 6.375000, 5.375000~
## $ trait_performance_avg <dbl> 5.000, 5.875, 5.250, 5.875, 5.500, 4.000, 4.750,~
```

Dealing with missing data

There are 44 columns in our dataset after cleaning. Although there are 216 rows with each row representing 1 participant's survey responses, many of these rows have lots of missing data. Therefore, I removed 92 participants data because they didn't answer any of the questions we used for analysis. After this, 124 participants remained. Most of them had no missing data, but there was 1 person with 1 missing response and another with 2 missing responses.




```
## [1] "Summary statistics for the age of removed participants:"
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   18.00   25.00   29.00   32.92   40.00   60.00       1
```

```
## [1] "For comparisons' sake, here are the summary statistics for the age of the included participants
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   22.75   30.00   34.61   44.50   72.00
```

```
## [1] "Gender of removed participants:"
```

```
##
## female   male
##     48     43
```

```
## [1] "Gender of remaining participants:"
```

```
##
## female   male
##     69     55
```

There is no clear reason why these people didn't finish the survey based on the little data we have about their age and gender. People across the age range dropped out and both males and females were about equally likely to drop out (about 41% of the females and 43% of the males didn't complete the survey).

Conditions

Here is a summary of our sample and how many people we had per condition

Var1	Freq
learning	63
performance	61

Var1	Freq
in	65
out	59

	learning	performance
in	35	30
out	28	31

Var1	Freq
female	69
male	55

	learning	performance
female	39	30
male	24	31

	in	out
female	40	29
male	25	30

Prompt received for learning priming:

Please think of a time when you met someone new in a social situation. This should be a time when you aimed to understand that person's values, attitudes or background. Please describe the situation in about 1-2 paragraphs. Also, please explain how attempting to learn about a stranger helped you extend your mastery of social situations.

Prompt received for performance priming:

Please think of a time when you met someone new in a social situation. This should be a time when you wanted to be judged favorably by that person. Please describe the situation in about 1-2 paragraphs. Also, please explain how conveying that you are likeable to a stranger helped you perform well in the social interaction.

In-group stranger paragraphs:

- If female: Who am I generally? I'm a 21 year old White woman, curly hair, athletic. I'm a student. I'd like to think I'm pretty nice and hardworking. I don't really know what else I'm supposed to say... I'm supposed to write about something interesting that's happened to me lately. Well, I don't know if this is interesting, but the only thing that I can seem to think of is that two days ago my boyfriend and I broke up. We've been dating since our junior year in high school and have been really close. It's been great being at college together. I thought he felt the same way, but I guess that things have changed. Now he wants to date other people. He says that he still cares a lot about me, but he doesn't want to be tied down to just one person. I've been kind of upset. It's all I think about. My friends all tell me that I'll meet other guys and all I need is for something good to happen to cheer me up. I guess they're right, but so far that hasn't happened.
- If male: Who am I generally? I'm a 21 year old White man, curly hair, athletic. I'm a student. I'd like to think I'm pretty nice and hardworking. I don't really know what else I'm supposed to say... I'm supposed to write about something interesting that's happened to me lately. Well, I don't know if this is interesting, but the only thing that I can seem to think of is that two days ago my girlfriend and I broke up. We've been dating since our junior year in high school and have been really close. It's been great being at college together. I thought she felt the same way, but I guess that things have changed. Now she wants to date other people. She says that she still cares a lot about me, but she doesn't want to be tied down to just one person. I've been kind of upset. It's all I think about. My friends all tell me that I'll meet other girls and all I need is for something good to happen to cheer me up. I guess they're right, but so far that hasn't happened.

Out-group stranger paragraphs:

- If female: Who am I generally? I'm a 21 year old Black woman, curly hair, athletic. I'm a student. I'd like to think I'm pretty nice and hardworking. I don't really know what else I'm supposed to say... I'm supposed to write about something interesting that's happened to me lately. Well, I don't know if this is interesting, but the only thing that I can seem to think of is that two days ago my boyfriend and I broke up. We've been dating since our junior year in high school and have been really close. It's been great being at college together. I thought he felt the same way, but I guess that things have changed. Now he wants to date other people. He says that he still cares a lot about me, but he doesn't want to be tied down to just one person. I've been kind of upset. It's all I think about. My friends all tell me that I'll meet other guys and all I need is for something good to happen to cheer me up. I guess they're right, but so far that hasn't happened.
- If male: Who am I generally? I'm a 21 year old Black man, curly hair, athletic. I'm a student. I'd like to think I'm pretty nice and hardworking. I don't really know what else I'm supposed to say... I'm supposed to write about something interesting that's happened to me lately. Well, I don't know if this is interesting, but the only thing that I can seem to think of is that two days ago my girlfriend and I broke up. We've been dating since our junior year in high school and have been really close. It's been great being at college together. I thought she felt the same way, but I guess that things have changed. Now she wants to date other people. She says that she still cares a lot about me, but she doesn't want to be tied down to just one person. I've been kind of upset. It's all I think about. My friends all tell me that I'll meet other girls and all I need is for something good to happen to cheer me up. I guess they're right, but so far that hasn't happened.

Quality checks

attention check:

To ensure that people were paying attention, they were asked to describe what they had read. Some people did not answer that question, so they may not have been paying attention

```
## [1] "How many people answered the first attention check, which asked them to summarize what they rea
```

Var1	Freq
TRUE	124

Everyone answered the first attention check, and everyone except for one person gave reasonable answers. This person was using vulgar language in their responses and was mad that the survey didn't pay more. I will be removing this person's responses.

Var1	Freq
0	120
1	4

Almost no one passed the second attention chedck, which asked them to pick a specific answer to prove they were paying attention. There may be some concerns about the quality of this data. People may have been paying attention in the beginning of the survey and faded as it went on. We can't remove people who failed the attention check since we would have no data left.

include or exclude:

Participants read this: The study is very important to us and a considerable amount of time and effort has gone into creating this survey. As such, if for whatever reason you feel that you did not complete the survey carefully or accurately, it would be extremely helpful if you could let us know this now. Your response will in no way affect your compensation or reputation on Mturk.

And then responded with:

- I DID NOT complete the survey carefully or accurately. Please exclude my responses from analysis.
- I DID complete the survey carefully or accurately. Please include my responses from analysis.

```
## [1] "How many people said to include or exclude their responses?"
```

```
##
## exclude include
##       2     121
```

Should we remove these 2 participants data?

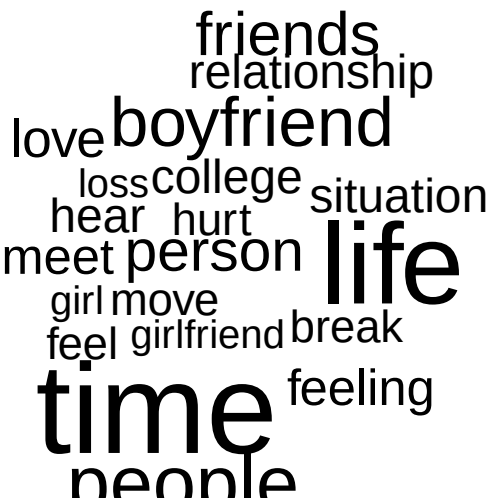
Textual analysis

Columns that we could perform textual analysis on and their prompts:

- `self_disc`: Now we would like to get to know a little bit more about you. Please write one paragraph about who you are generally (keep it anonymous), and please describe something interesting that has happened in your life recently. This paragraph will randomly be shown to future Mturk participants to read.
- `note`: Now you have the opportunity to write a note to the participant you just read about. Please respond with your thoughts and feelings to his/her "interesting thing that has happened". Your message will be emailed to this participant (and he/she will not receive any other information about you).Please write a 1-2 paragraph response below:

Top 20 Words for In−Group Participants

Top 20 Words for Out−Group Participants




```
## I calculated the average sentiment expressed per person. Sometimes none of the words that a person us

## [1] "Here is the final data frame:"

## Rows: 124
## Columns: 48
## $ participantID              <chr> "1", "2", "3", "4", "5", "6", "7", "8", "9~
## $ startDate                  <chr> "5/22/13 16:54", "5/23/13 16:15", "5/23/13~
## $ endDate                    <chr> "5/22/13 17:08", "5/23/13 16:20", "5/23/13~
## $ age                        <chr> "19", "26", "28", "38", "18", "21", "20", ~
## $ gender                     <fct> female, female, male, male, male, male, fe~
## $ self_disc                  <chr> "I'm a generally happy person and I like t~
## $ priming_response           <chr> "We had a date night for my sorority a cou~
## $ att_check                  <chr> "She and her boyfriend broke up", "The bre~
## $ note                       <chr> "My boyfriend and I broke up a little over~
## $ pt_1                       <dbl> 1, 2, 2, 2, 3, 4, 1, 1, 1, 1, 2, 1, 1, 2, ~
## $ pt_2                       <dbl> 7, 6, 6, 6, 5, 4, 7, 7, 6, 6, 6, 7, 7, 5, ~
## $ pt_3                       <dbl> 2, 1, 2, 1, 4, 4, 1, 1, 1, 1, 3, 1, 1, 4, ~
## $ pt_4                       <dbl> 7, 5, 6, 7, 5, 4, 7, 6, 7, 7, 4, 7, 6, 3, ~
## $ pt_5                       <dbl> 1, 2, 2, 1, 2, 2, 1, 1, 1, 1, 6, 1, 1, 1, ~
## $ emp                        <chr> "6", "5", "5", "6", "3", "4", "7", "7", "7~
## $ dg                         <dbl> 0, 5, 0, 20, 0, 5, 2, 12, 0, 25, 0, 0, 10,~
## $ lp_scale_1_1               <dbl> 7, 6, 4, 6, 6, 4, 7, 6, 6, 6, 3, 6, 5, 4, ~
## $ lp_scale_1_2               <dbl> 4, 7, 4, 1, 5, 4, 3, 6, 5, 7, 3, 7, 5, 4, ~
## $ lp_scale_1_3               <dbl> 5, 6, 6, 5, 5, 4, 6, 6, 7, 6, 5, 5, 6, 6, ~
## $ lp_scale_1_4               <dbl> 6, 5, 6, 7, 4, 4, 7, 6, 6, 7, 6, 6, 6, 5, ~
## $ lp_scale_1_5               <dbl> 5, 5, 6, 7, 5, 4, 7, 6, 7, 7, 6, 4, 5, 5, ~
## $ lp_scale_1_6               <dbl> 7, 7, 4, 6, 3, 4, 3, 7, 5, 7, 6, 5, 5, 4, ~
## $ lp_scale_1_7               <dbl> 4, 4, 6, 6, 6, 4, 3, 5, 6, 7, 3, 5, 5, 2, ~
## $ lp_scale_1_8               <dbl> 7, 5, 6, 6, 7, 4, 7, 6, 7, 6, 6, 6, 6, 6, ~
## $ lp_scale_1_9               <dbl> 3, 5, 6, 7, 4, 4, 6, 6, 7, 6, 6, 5, 5, 5, ~
## $ lp_scale_2_1               <dbl> 6, 6, 6, 7, 7, 4, 7, 6, 6, 6, 5, 7, 7, 4, ~
## $ lp_scale_2_2               <dbl> 6, 7, 6, 7, 6, 4, 7, 6, 7, 7, 2, 7, 6, 4, ~
## $ lp_scale_2_3               <dbl> 4, 5, 6, 7, 6, 4, 4, 2, 6, 6, 3, 6, 5, 3, ~
## $ lp_scale_2_4               <dbl> 5, 6, 6, 7, 5, 4, 5, 6, 6, 6, 4, 7, 5, 5, ~
```

```
## $ lp_scale_2_5                <dbl> 7, 7, 6, 6, 7, 4, 7, 6, 7, 6, 5, 6, 6, 5, ~
## $ att_check2                  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ lp_scale_2_7                <dbl> 3, 6, 6, 7, 6, 4, 6, 2, 5, 5, 5, 7, 6, 4, ~
## $ lp_scale_2_8                <dbl> 6, 5, 4, 6, 5, 7, 6, 6, 7, 5, 7, 6, 6, 5, ~
## $ includeOrExclude            <fct> NA, include, include, include, include, ex~
## $ priming_received            <chr> "performance", "learning", "performance", ~
## $ inOrOut                     <chr> "in", "out", "out", "out", "in", "out", "i~
## $ condition                   <chr> "perf_in", "learn_out", "learn_out", "lear~
## $ pt_1_reversed               <dbl> 7, 6, 6, 6, 5, 4, 7, 7, 7, 7, 6, 7, 7, 6, ~
## $ pt_3_reversed               <dbl> 6, 7, 6, 7, 4, 4, 7, 7, 7, 7, 5, 7, 7, 4, ~
## $ pt_5_reversed               <dbl> 7, 6, 6, 7, 6, 6, 7, 7, 7, 7, 2, 7, 7, 7, ~
## $ PT_avg                      <dbl> 6.8, 6.0, 6.0, 6.6, 5.0, 4.4, 7.0, 6.8, 6.~
## $ trait_learning_avg          <dbl> 5.571429, 5.625000, 5.750000, 6.375000, 5.~
## $ trait_performance_avg       <dbl> 5.000, 5.875, 5.250, 5.875, 5.500, 4.000, ~
## $ avg_sentiment               <dbl> 0.1666667, 1.0000000, 3.0000000, 0.6666667~
## $ words_with_sentiment        <dbl> 6, 1, 1, 6, 0, 3, 7, 5, 10, 10, 0, 3, 4, 4~
## $ first_person_pronoun_count  <dbl> 13, 2, 1, 2, 1, 0, 20, 2, 5, 1, 0, 8, 5, 2~
## $ second_person_pronoun_count <dbl> 6, 5, 0, 11, 0, 4, 8, 3, 8, 13, 1, 4, 12, ~
## $ third_person_pronoun_count  <dbl> 2, 0, 3, 3, 0, 4, 2, 1, 9, 3, 0, 2, 3, 0, ~
```