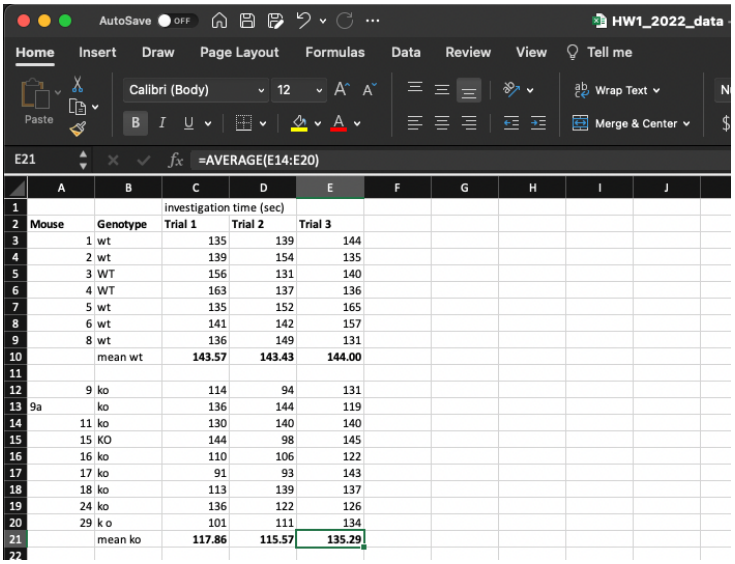
Neural Data Science Spring 2022
Homework 1

You have an Excel sheet, “HW1_2022_data.xlsx”, that contains some data of the kind that you might receive for analysis.



	A	B	C	D	E	F	G	H	I	J
1			investigation time (sec)							
2	Mouse	Genotype	Trial 1	Trial 2	Trial 3					
3	1	wt	135	139	144					
4	2	wt	139	154	135					
5	3	WT	156	131	140					
6	4	WT	163	137	136					
7	5	wt	135	152	165					
8	6	wt	141	142	157					
9	8	wt	136	149	131					
10		mean wt	**143.57**	**143.43**	**144.00**					
11										
12	9	ko	114	94	131					
13	9a	ko	136	144	119					
14	11	ko	130	140	140					
15	15	KO	144	98	145					
16	16	ko	110	106	122					
17	17	ko	91	93	143					
18	18	ko	113	139	137					
19	24	ko	136	122	126					
20	29	k o	101	111	134					
21		mean ko	**117.86**	**115.57**	**135.29**					
22										

Your assignment is to 1) read this data into R, 2) put it into “long” format for analysis, and then 3) obtain summary statistics of the mean and standard deviation for investigation time for each trial, grouped by genotype. The data should look like this when you are finished reading it in and wrangling it:

```
# A tibble: 48 × 4
   Mouse Genotype  Trial Investigation_Time
   <chr> <fct>     <fct>              <dbl>
 1 1     wild_type 1                    135
 2 1     wild_type 2                    139
 3 1     wild_type 3                    144
 4 2     wild_type 1                    139
 5 2     wild_type 2                    154
 6 2     wild_type 3                    135
 7 3     wild_type 1                    156
 8 3     wild_type 2                    131
 9 3     wild_type 3                    140
10 4     wild_type 1                    163
11 4     wild_type 2                    137
12 4     wild_type 3                    136
13 5     wild_type 1                    135
14 5     wild_type 2                    152
15 5     wild_type 3                    165
16 6     wild_type 1                    141
17 6     wild_type 2                    142
18 6     wild_type 3                    157
19 8     wild_type 1                    136
20 8     wild_type 2                    149
21 8     wild_type 3                    131
22 9     knockout  1                    114
23 9     knockout  2                     94
24 9     knockout  3                    131
25 9a    knockout  1                    136
26 9a    knockout  2                    144
27 9a    knockout  3                    119
28 11    knockout  1                    130
29 11    knockout  2                    140
30 11    knockout  3                    140
31 15    knockout  1                    144
32 15    knockout  2                     98
33 15    knockout  3                    145
34 16    knockout  1                    110
35 16    knockout  2                    106
36 16    knockout  3                    122
37 17    knockout  1                     91
38 17    knockout  2                     93
39 17    knockout  3                    143
40 18    knockout  1                    113
41 18    knockout  2                    139
42 18    knockout  3                    137
43 24    knockout  1                    136
44 24    knockout  2                    122
45 24    knockout  3                    126
46 29    knockout  1                    101
47 29    knockout  2                    111
48 29    knockout  3                    134
```

You should submit (via Slack DM) your R code before the beginning of class **January 31**. Do not modify the Excel file: your code should run on the original file provided. The final lines of the code should print out the data and then generate the summary statistics. The “tidyverse” and “readxl” libraries will be extremely helpful. (Hint: check out "stringr", included in tidyverse.)