Question 1

A. For the two experimental conditions (steering focus or "dualSteerFocus"; dialing focus or "dualDialFocus") report the mean (M) of the total dialing time, standard deviation (SD), and standard error of the mean (SE) of total dialing time.

	M	SD	SE
Dialing focus	4.0	1.6	0.5
Steering focus	5.9	2.3	0.7

The mean (M), standard deviation (SD) and standard error of the mean (SE) is rounded given the precision is 1 digit after the decimal point.

The total dialing time is calculated from `phoneNrLengthAfterKeyPress == 0` to `phoneNrLengthAfterKeyPress == 12` (instead of 11 as written in the hints), and the reason is based on the following facts:

- When `phoneNrLengthAfterKeyPress == 11`, `PhoneNrSoFar` is 785432569; whereas when `phoneNrLengthAfterKeyPress == 12`, `PhoneNrSoFar` is 7854325698 ("8" is appended at the tail).

B. Do the same for average absolute lateral deviation (in m).

	M	SD	SE
Dialing focus	0.4669684	0.4389860	0.3104099
Steering focus	0.6740285	0.6854105	0.4846584

Below is the absolute lateral deviation calculated per each keypress in two experimental conditions respectively.

	M	SD	SE
Dialing focus	0.09507108	0.1526213	0.04405799
Steering focus	0.09689894	0.1465787	0.04231361

To illustrate, the source data frame used for the calculation is as below:

partOfExperiment	pp	trial	phoneNrLengthAfterKeyPress	lanePosition	diff
dualSteerFocus	1	7	0	-0.101	0.000
dualSteerFocus	1	7	1	-0.101	0.000
dualSteerFocus	1	7	2	-0.104	0.003
dualSteerFocus	1	7	3	-0.110	0.006
dualSteerFocus	1	7	4	-0.115	0.005
dualSteerFocus	1	7	5	-0.119	0.004
dualSteerFocus	1	7	6	-0.124	0.005
dualSteerFocus	1	7	7	-0.130	0.006
dualSteerFocus	1	7	8	-0.135	0.005
dualSteerFocus	1	7	9	-0.141	0.006

dualSteerFocus	1	7	10	-0.146	0.005
dualSteerFocus	1	7	11	-0.152	0.006
dualSteerFocus	1	7	12	-0.162	0.010

"diff" is the absolute difference value between each keypress, based on which the mean (M), standard deviation (SD) and standard error of the mean (SE) is calculated.

As there is no movement (so no "diff") at the starting point where `phoneNrLengthAfterKeyPress == 0`, this row is removed while calculating the figures in the above table.

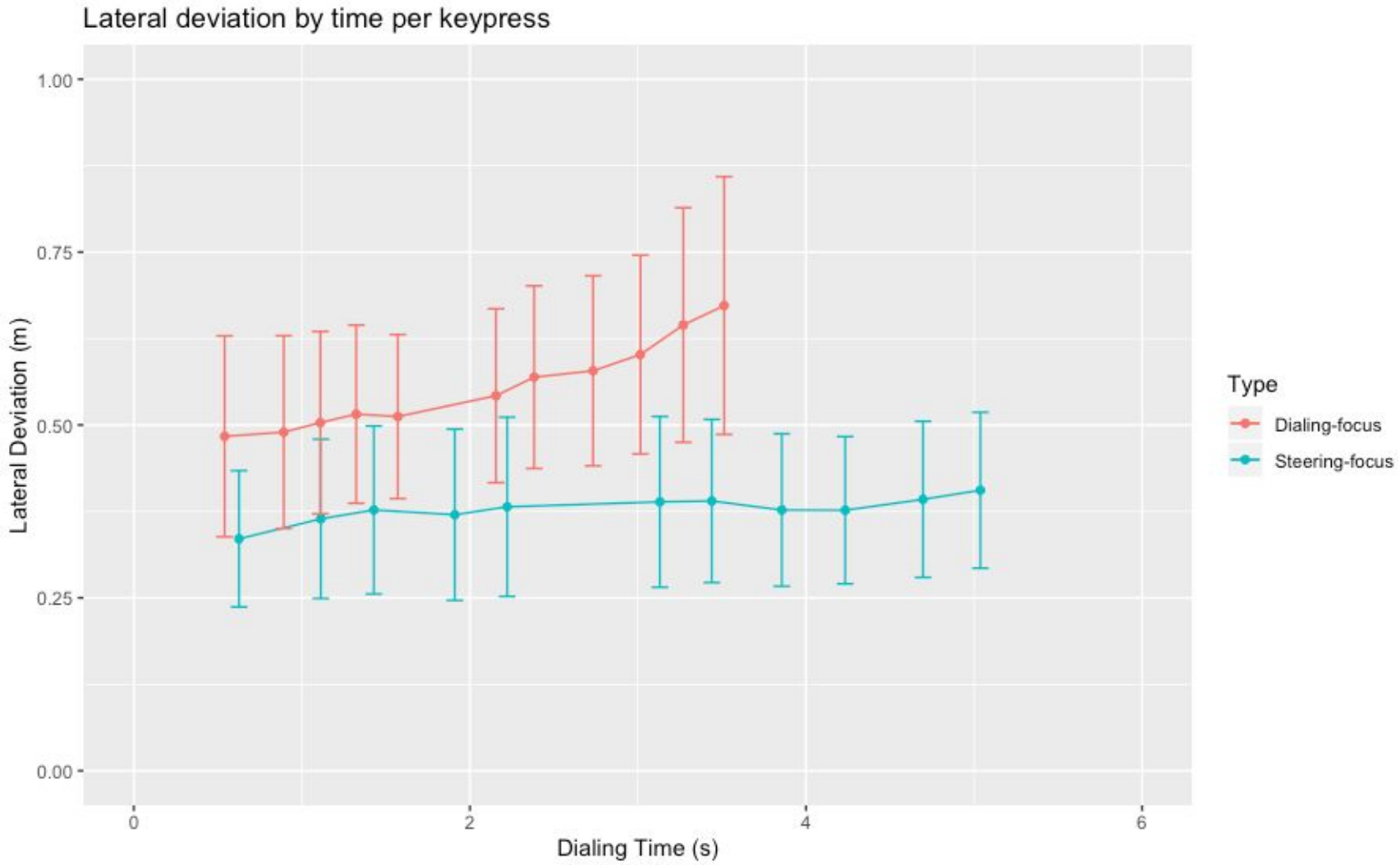
Below is the absolute lateral deviation calculated per the whole dialing task in two experimental conditions respectively.

	M	SD	SE
Dialing focus	0.6558922	0.7759643	0.2240016
Steering focus	0.3882766	0.4262150	0.1230377

To illustrate, the source data frame used for the calculation is as below:

partOfExperiment	pp	trial	phoneNrLengthAfterKeyPress	lanePosition	diff
dualSteerFocus	1	7	0	-0.101	0.000
dualSteerFocus	1	7	12	-0.162	0.061

C. Make a plot of how lane deviation changes over time with each keypress.



The source data frame is filtered from the observations where `phoneNrLengthAfterKeyPress` is between 1 and 11, the experimental conditions are either `dualSteerFocus` or `dualDialFocus` and there are no typing errors made on trial.

Afterwards the source data frame is grouped by `phoneNrLengthAfterKeyPress` and `partOfExperiment` before calculating the mean of `lanePosition` and `timeRelativeToTrialStart`.

For the error bars, the upper and lower limit is calculated using the standard error of the lateral deviation, and as reflected in the diagram above, the standard error varies in a wide range. The potential reason might be as below:

- At the starting point, the lane position varies a lot. e.g., for participant 3 in trial 27 for the dialing focus task, the lane position is 0.076, whereas for participant 4 in trial 19 for the same task, the lane position is 0.790.

D. Based on the plot in question C, answer the following question: "For each condition argue whether you think the average participant waited until the natural breakpoint (between 5th and 6th digit) with interleaving dialing for driving." Motivate your opinion clearly based on the observed pattern.

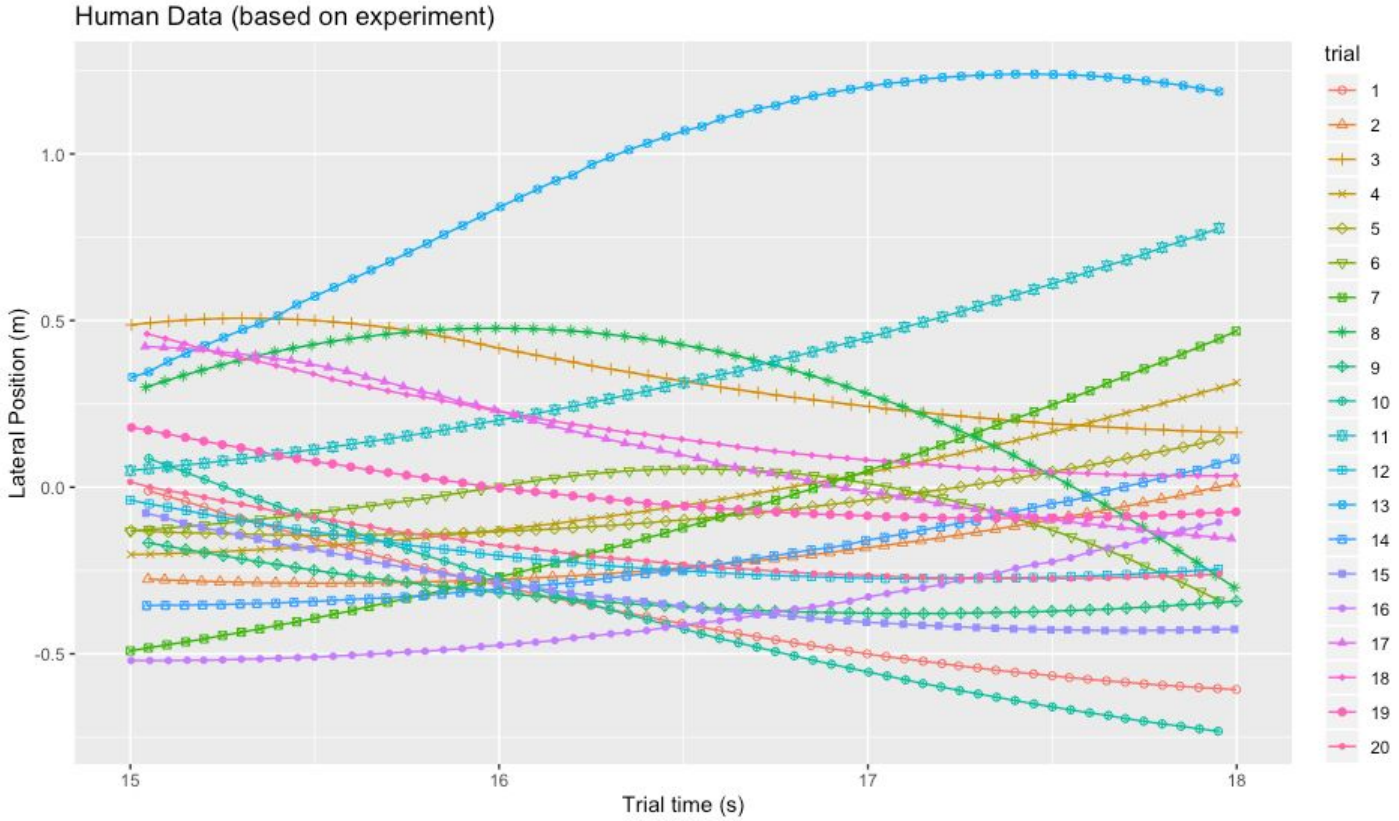
Only looking at the diagram above we can see the following patterns:

- For both conditions, the time between keypress five and keypress six is longer than the time between other consecutive keypresses.

As the participants are tested with dialing a UK-style telephone number (i.e., 07854-325698), and the diagram above displays the average time between keypresses for all participants per each experimental condition, it can be deduced that the participants on average wait after a natural subtask boundary (that is, until the natural breakpoint) before resuming the dialing task again. Considering the above telephone number 07854-325698 as an example, it takes longer time between 07854 and 325698 than any other consecutive numbers.

Question 2

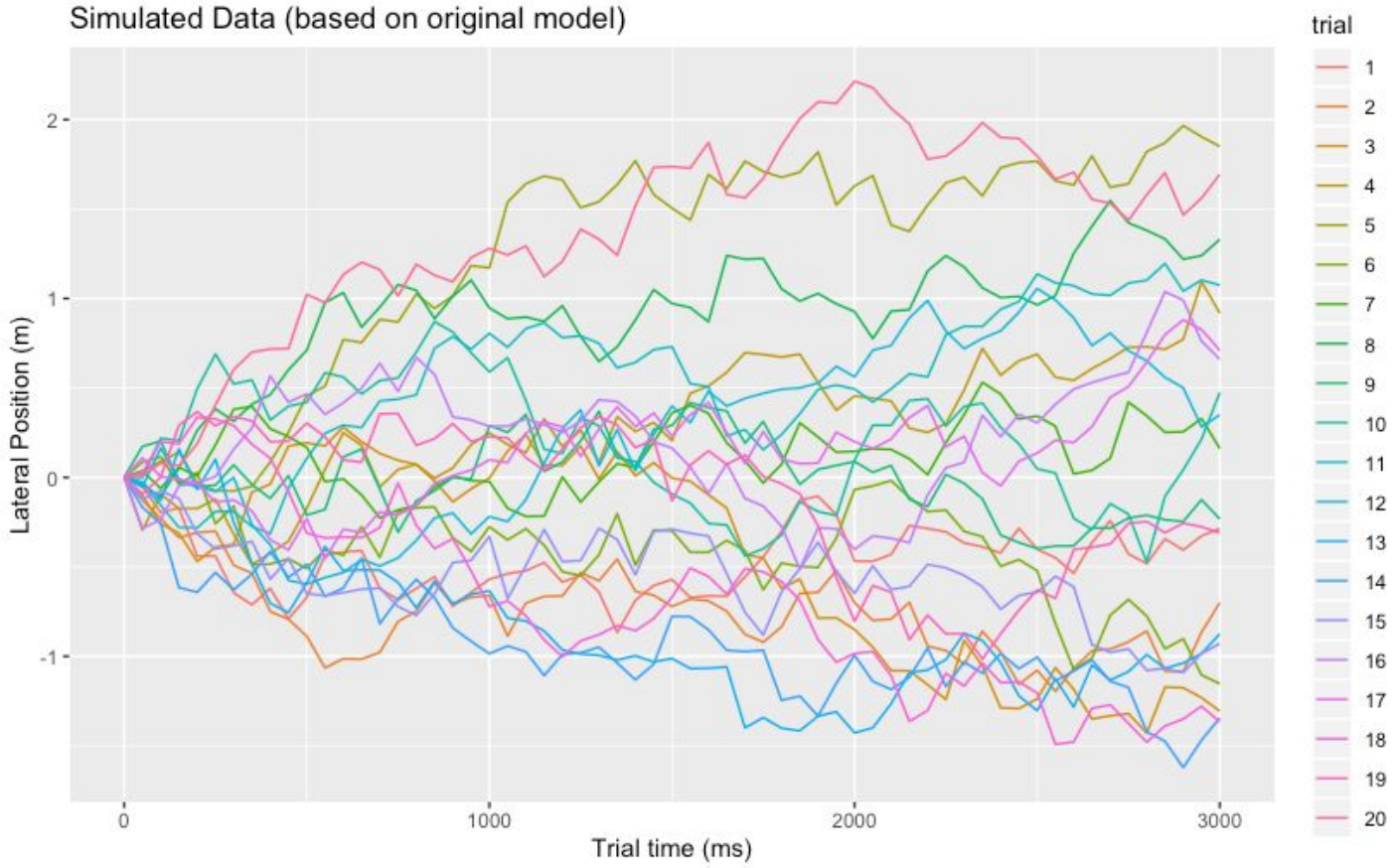
A. Create a plot in which you show how the lateral position (posX) of the car changes over time (trialTime) for the subset of data between 15000 and 18000 milliseconds (i.e., between 15 and 18 seconds). Make sure that the data for each trial has a unique color.



The observations are filtered by `trialTime` between 15s and 18s, which are then grouped by `trial`.

As the color of the line for each trial is a little bit indistinguishable, the shape and color of points are added for us to better differentiate between each trial.

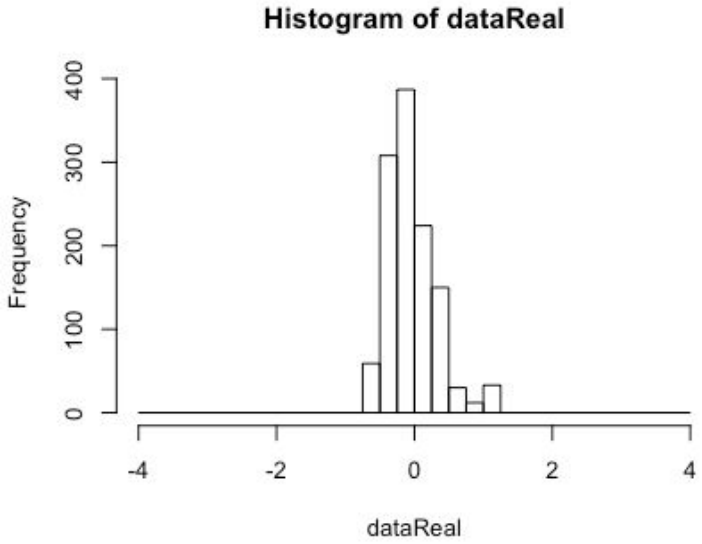
B. Generate simulated lateral position data for (at least) 20 simulated trials in which you assume that the car starts drifting from a point 0, and samples values from the modeling distribution (M = 0, SD = 0.13) every 50 milliseconds for a period of 3000 milliseconds.

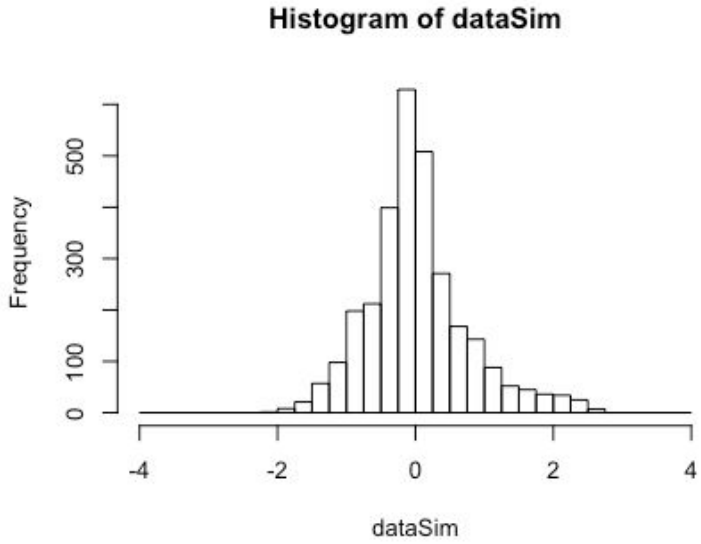


The lateral position is calculated by cumsum() based on rnorm(), and part of the data frame for this diagram is as below:

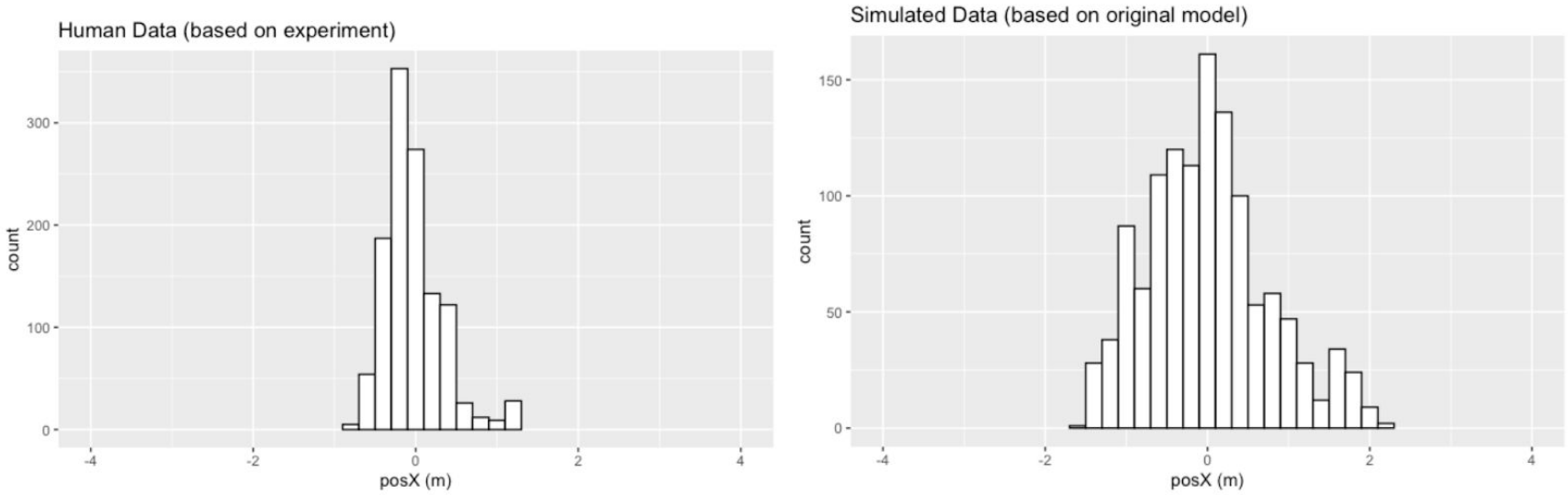
trial	trialTime	posX
1	0	0.00000000
1	50	-0.05956592
1	100	-0.05270244
1	...	...
1	3000	-1.45266936

C. You will probably notice that the shape of the simulated lines is very different from that of the human data. To see this even more clearly, I want you to create two histograms for your report.





The histograms above are drawn with the x-axis equal to `posX` and binwidth equal to 0.2 within the range [-4, 4].



D. For the two histograms you plotted in question C you can also calculate the standard deviation of each of the two datasets.

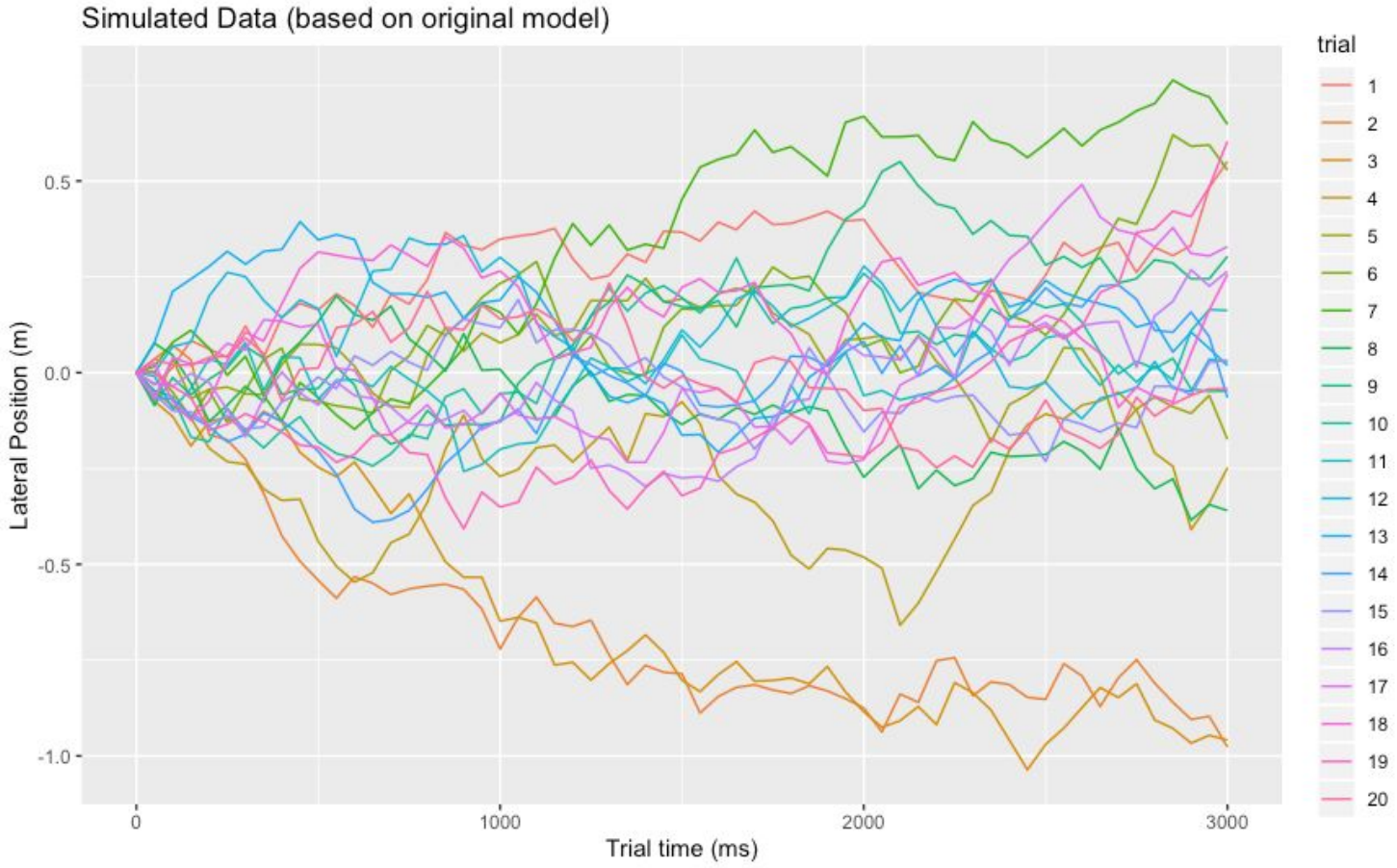
Type	SD
Empirical human data	0.3596126
Simulated model data	0.735189/0.6149442

SD of simulated model data is larger than SD of empirical human data, as it can also be illustrated in the histogram above that the span of simulated model data is wider than that of empirical human data across the range [-4, 4].

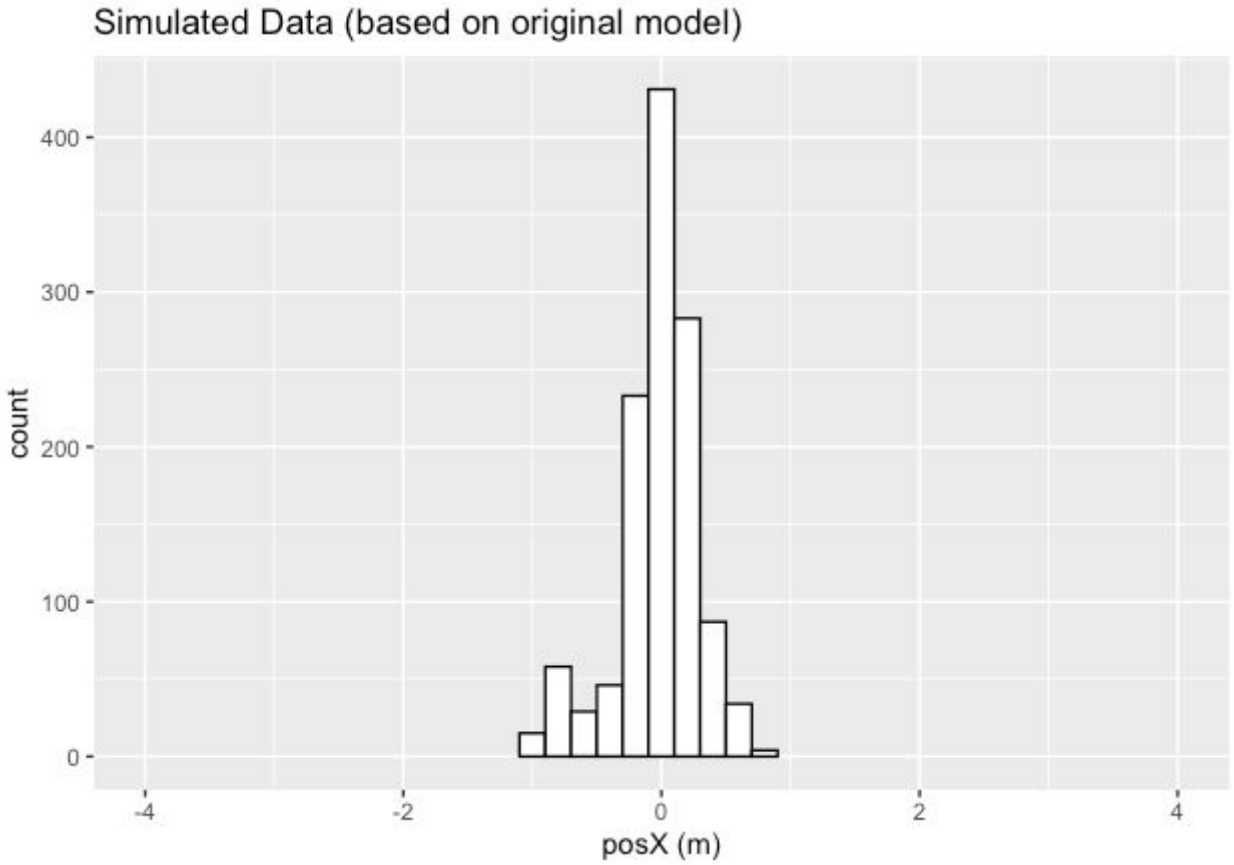
E. Try different values in this model (instead of SD = 0.13), and decide which value results in a distribution that is closest to the human data.

(i) The final SD to set the model is 0.06.

(ii) A plot of how lane position changes over time for the individual simulated trials is as below:



(iii) A plot of the resulting distribution is as below:



(iv) The SD of this resulting distribution is 0.35/0.31.

Question 3

A. What was the average value of the interkeypress intervals?

The average value of the interkeypress intervals is calculated by taking the data by `partOfExperiment == "singleDialing2" & typingErrorMadeOnTrial == 0`.

Moreover, the data is also filtered to remove the noise data as below, as they are partial incomplete data without typing the whole telephone number. So it can be counted as a kind of typing error.

partOfExperiment	pp	trial	PhoneNrSoFar	timeRelativeToTrialStart	phoneNrLengthAfterKeyPress
singleDialing2	2	32		0	0
singleDialing2	8	41		0	0
singleDialing2	8	41		338	1
singleDialing2	8	41		656	2
singleDialing2	10	48		0	0
singleDialing2	12	41	0	0	0

Afterwards the data is pre-processed with the difference of `timeRelativeToTrialStart` calculated between each keypress as below:

pp	trial	phoneNrLengthAfterKeyPress	timeRelativeToTrialStart	diff
1	1	0	0	0

1	1	1	9	9
1	1	2	18	9
1	1	3	26	8
1	1	4	35	9
1	1	5	198	163
1	1	6	431	233
1	1	7	597	166
1	1	8	792	195
1	1	9	946	154
1	1	10	1128	182
1	1	11	1347	219
1	1	12	1546	199

Then the mean is calculated based on `diff` by grouping each participant as below:

pp	mean
1	172.1410
2	297.0192
3	279.6462

4	276.4462
5	261.0923
6	272.1231
7	220.8000
8	255.6154
9	291.1346
10	237.2923
11	250.6154
12	220.6538

Finally the mean for “the average interkeypress interval for the average participant” is calculated with mean() on all participants, with the round function as below:

```
meanKpi <- round(meanKpi / 10) * 10
```

The result was an average of **250ms.**

B. What value did you pick for your model and why?

The values we pick for our model are as below:

- `gaussDeviateSD <- 0.35`
- `gaussDriveNoiseSD <- 0.1`
- `singleTaskKeyPressTimes <- rep(250, 11)`

The reasons are as below:

- We want our model to closely simulate the human behavior, so `gaussDeviateSD` is chosen to be the SD from the distribution which is closest to the scenarios when human drives a car while typically avoiding looking at the road for at most 3 to 4 seconds.
- As `gaussDriveNoiseSD` is the value when the driver is actively controlling the car, so `gaussDriveNoiseSD` is chosen to be smaller than `gaussDeviateSD.`
- `For singleTaskKeyPressTimes`, we also want this value to closely imitate the human behaviour, so the average interkeypress interval for the average participant from the human data is used.