Data Exploration

Monica Chan

2021-10-08

# 1 Data Exploration

Complete the data loading/cleaning/wrangling part, as well as having a mostly complete data exploration part.

## 1.1 Data Loading

*What the data is*

Compilation of RT-PCR results for a couple of probes (N1 & N2) to detect recovered heat-inactivated SARS-CoV-2 samples that have undergone sample enhancement procedures. These data can be found in the data folder under raw\_data

## 1.2 Data Cleaning

Cleaned data is found under the data in processed\_data folder

1. Compiled reports into a single csv.
2. Import the reports into R.
3. Check summary, head, and tail of data
4. Select relevant parts of the data to form new data frame
5. Mutate classes of data for easier manipulation.
6. Renamed a value as it is the only value left. The Raw data had sevearl versions of it as the series of columns displayed a train of thought display of calculation.

* noted that there were additional NAs that would be generated when mutating Ct values into numeric. Which should be fine, but with how the reports are generated some levels recorded are noted as Undetermined or not reliably detected.
* Decide to Split the Data and make a companion UndeterminedData dataframe for additional analysis.

**Making the UndeterminedData** see above but added filter() for Ct==“Undetermined” to pull all data with Undetermined as a reported value

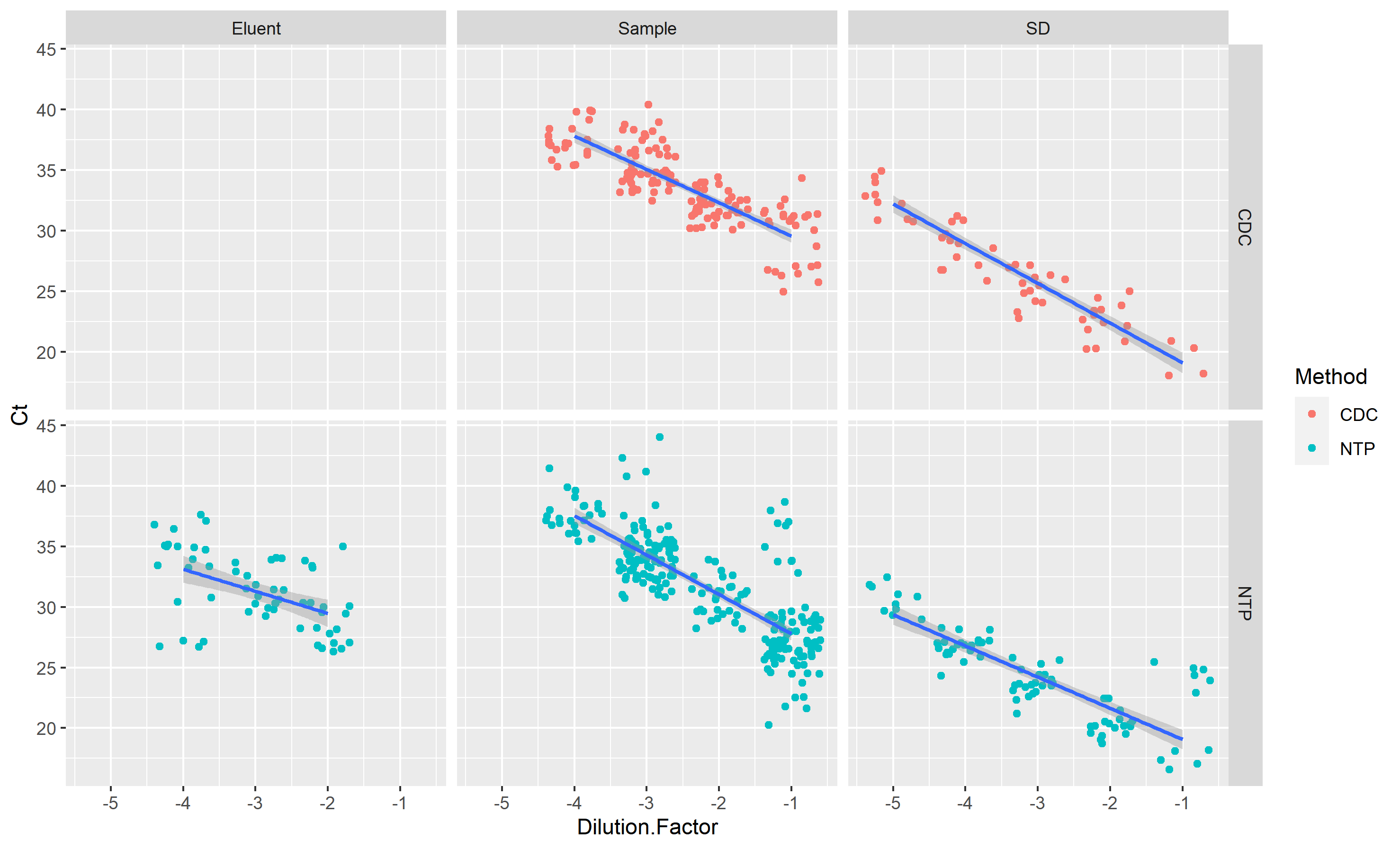
## 1.3 Data Wrangling

Started to play with analysisscript. Outputs of the data play are in results.

####Run summaries

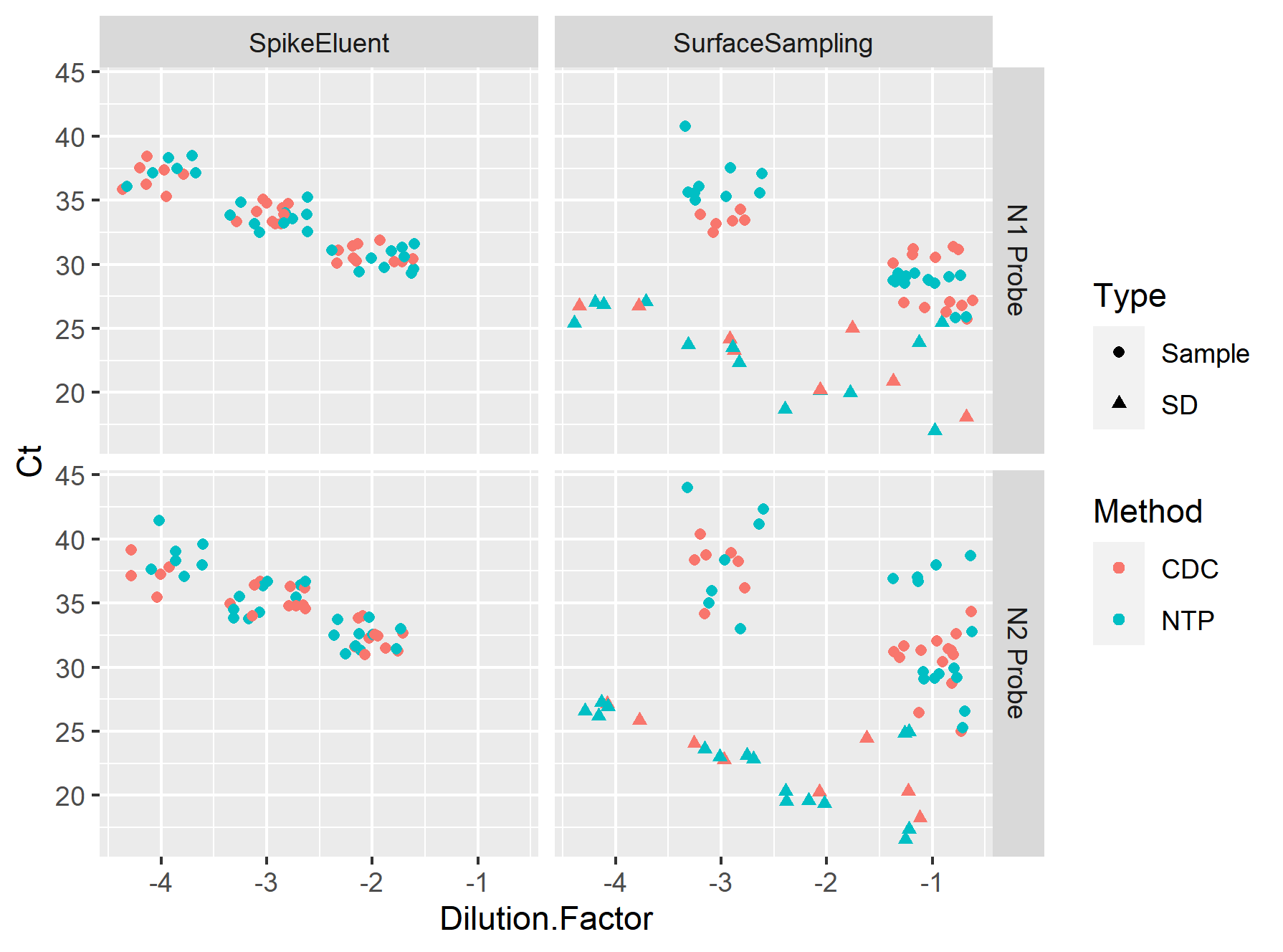
####Exploritory Figures

##### 1.3.0.0.1 View the whole data set



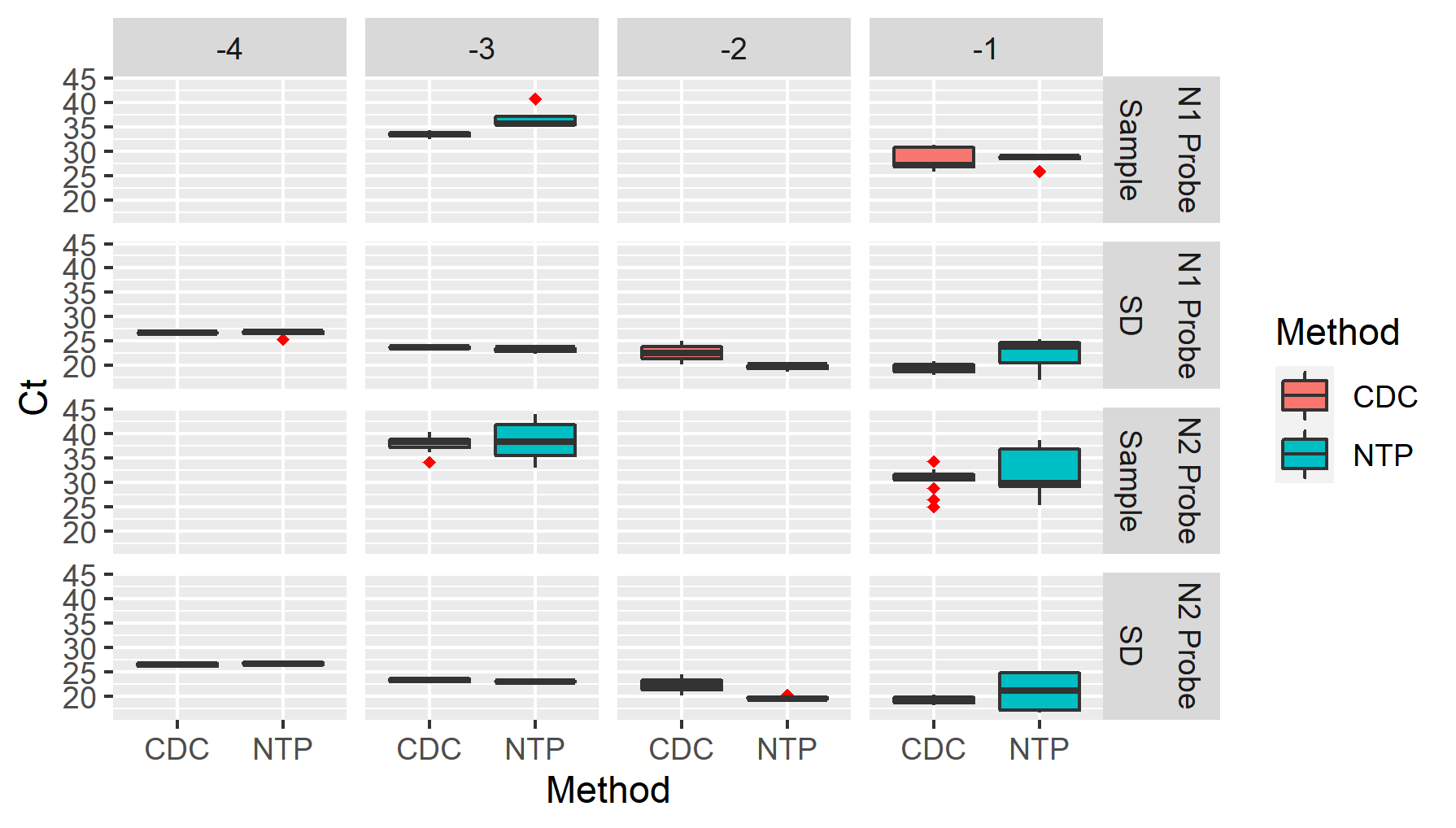
resultfigure

##### 1.3.0.0.2 Breakdown comparisons with the two methods being compared– CDC and NTP had a comparitive input of 300uL

 Interesting– why is there such a group difference in Surface Sampling experiements versus Spike Oh makes sense it’s SD- the serial dilution. We expect the SD to be like the standard curve and the recovery should have loss.

Deeper dive into looking at the Surface Sampling Experiment, as it has some differences and is new-er data. Also includes data with Temp & RH:

##### 1.3.0.0.3 Breakdown comparisons more with only the Surface Sample experiment



SS300