

# t\_test\_results\_reproduced

Andrew Willems and Tian Hong

3/22/2022

## T-test Analysis

The goal of this analysis is to replicate the t-test results seen in previous analysis

**Step One: Load needed packages.** Magrittr is a package that allows us to use pipes (`%>%`) in our code. Rstatix is used to do the pairwise t-tests and p-value correction. Tidyverse is used for data manipulation.

```
#Loading needed packages
library(magrittr)
library(rstatix)

##
## Attaching package: 'rstatix'

## The following object is masked from 'package:stats':
##
##   filter

library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5    v purrr  0.3.4
## v tibble  3.1.6    v dplyr  1.0.8
## v tidyr   1.2.0    v stringr 1.4.0
## v readr   2.1.2    v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x tidyr::extract() masks magrittr::extract()
## x dplyr::filter()  masks rstatix::filter(), stats::filter()
## x dplyr::lag()     masks stats::lag()
## x purrr::set_names() masks magrittr::set_names()
```

**Step Two: Load the data and make separate data frames that are comprised of only left or right hemisphere data.**

```
#Loading the data
setwd("~/Documents/PhD Program/Hong Lab/Projects/Neuron_Project/Data/")
pnn_data <- read.csv("processed_data_PNN_3cohorts.csv", sep = "\t")
pnn_data_left <- filter(pnn_data, Hemisphere=="Left")
pnn_data_right <- filter(pnn_data, Hemisphere=="Right")
```

**Step Three:** Look at how many of each condition we have in our data. If we don't have very many we will need to use a common standard deviation in our t-test. If the sample size is large enough we can set the common standard deviation to FALSE in t-test.

```
#Make table to look at how many of each condition we have to know if we need to
#use pooled standard deviation in our t-test
table(pnn_data_left$Condition)
```

```
##
## NH NW SH SW
## 55 52 56 59
```

```
table(pnn_data_right$Condition)
```

```
##
## NH NW SH SW
## 51 57 62 57
```

**Step Four:** Performing a simple calculation to determine if we need to do p-value

adjustment or not because of running multiple tests. For reference on formula see Goldman tutorial doc (pdf included in slack) and Jafari et al. (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6099145/>)

```
#Performing a simple calculation to determine if we need to do p-value
#adjustment or not because of running multiple tests. For reference on formula
#see Goldman tutorial doc (pdf included in slack) and Jafari et al.
#(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6099145/)
```

```
#We have 4 condition (NH, NW, SH, SW)
rate_of_false_positive <- 1-(1 - 0.05)^4
print(paste0("The false positive rate for our data is: ",
             round(rate_of_false_positive, digits = 4) *100, "%"))
```

```
## [1] "The false positive rate for our data is: 18.55%"
```

**Step Five:** Do pairwise t-test for left hemisphere. Our comparison is condition (NH, SH, NW, SW) and how they relate to Mean 2. We set pool.sd to FALSE because of large enough numbers for each of our groups. We adjust the p-value with the false discovery rate (FDR) because of a nearly 20% false positive rate that was calculated above. We set detailed to TRUE so that we get the most verbose output we can.

```
#Mean 2 left pairwise t-test
pwc_left_mean2 <- pnn_data_left %>%
  pairwise_t_test(
    Mean.2 ~ Condition, pool.sd = FALSE,
    p.adjust.method = "fdr",
    detailed = TRUE
  )
```

```
pwc_left_mean2
```

```
## # A tibble: 6 x 17
##   estimate estimate1 estimate2 .y. group1 group2  n1    n2 statistic      p
```

```
## *      <dbl>      <dbl>      <dbl> <chr> <chr> <chr> <int> <int>      <dbl> <dbl>
## 1      0.881      69.6      68.7 Mean~ NH      NW      55    52      0.483 6.3 e-1
## 2     -6.04      69.6      75.7 Mean~ NH      SH      55    56     -2.88 5 e-3
## 3      4.00      69.6      65.6 Mean~ NH      SW      55    59      2.19 3.1 e-2
## 4     -6.92      68.7      75.7 Mean~ NW      SH      52    56     -4.06 9.9 e-5
## 5      3.12      68.7      65.6 Mean~ NW      SW      52    59      2.31 2.3 e-2
## 6     10.0      75.7      65.6 Mean~ SH      SW      56    59      5.89 5.37e-8
## # ... with 7 more variables: df <dbl>, conf.low <dbl>, conf.high <dbl>,
## #   method <chr>, alternative <chr>, p.adj <dbl>, p.adj.signif <chr>
```

Do same as above but for right hemisphere data.

```
#Mean 2 right pairwise t-test
pwc_right_mean2 <- pnn_data_right %>%
  pairwise_t_test(
    Mean.2 ~ Condition, pool.sd = FALSE,
    p.adjust.method = "fdr",
    detailed = TRUE
  )

pwc_right_mean2

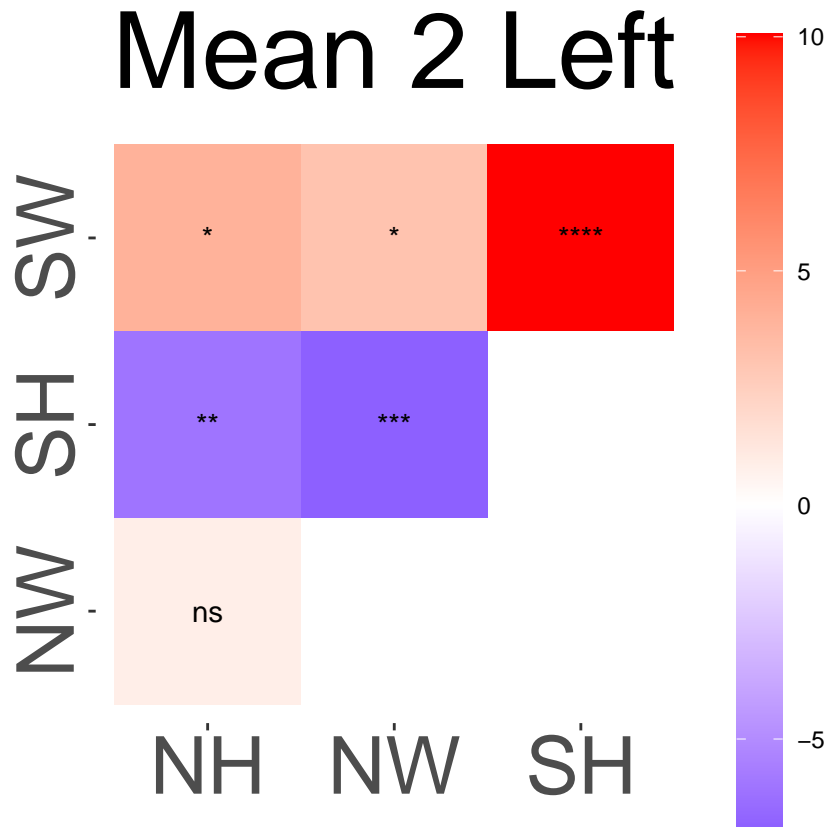
## # A tibble: 6 x 17
##   estimate estimate1 estimate2 .y. group1 group2 n1 n2 statistic p
## *      <dbl>      <dbl>      <dbl> <chr> <chr> <chr> <int> <int>      <dbl> <dbl>
## 1      0.612      69.8      69.2 Mean~ NH      NW      51    57      0.346 7.3 e-1
## 2     -4.54      69.8      74.3 Mean~ NH      SH      51    62     -2.36 2 e-2
## 3      3.34      69.8      66.5 Mean~ NH      SW      51    57      1.93 5.7 e-2
## 4     -5.15      69.2      74.3 Mean~ NW      SH      57    62     -3.36 1 e-3
## 5      2.72      69.2      66.5 Mean~ NW      SW      57    57      2.13 3.5 e-2
## 6      7.88      74.3      66.5 Mean~ SH      SW      62    57      5.29 6.32e-7
## # ... with 7 more variables: df <dbl>, conf.low <dbl>, conf.high <dbl>,
## #   method <chr>, alternative <chr>, p.adj <dbl>, p.adj.signif <chr>
```

Step Six: Plotting the heatmap of the left hemisphere data.

```
#Heatmap for Mean 2 left hemisphere
mid <- 0

ggplot(data = pwc_left_mean2, aes(x=group1, y=group2, fill=estimate))+
  geom_tile()+
  theme(plot.title = element_text(size = 40, face = "plain", hjust = 0.5),
        panel.background = element_blank(),
        axis.text.y = element_text(size = 30, face = "plain", angle = 90,
                                     hjust = 0.5),
        axis.text.x = element_text(size = 30, face = "plain"),
        legend.title = element_blank(),
        legend.key.height = unit(60, units = "pt"))+
  ggtitle("Mean 2 Left")+
  scale_fill_gradient2(low = "blue", high = "red", midpoint = mid,
                      mid = "white")+
  xlab("")+
  ylab("")+
```

```
coord_equal()+
geom_text(aes(label=p.adj.signif))
```

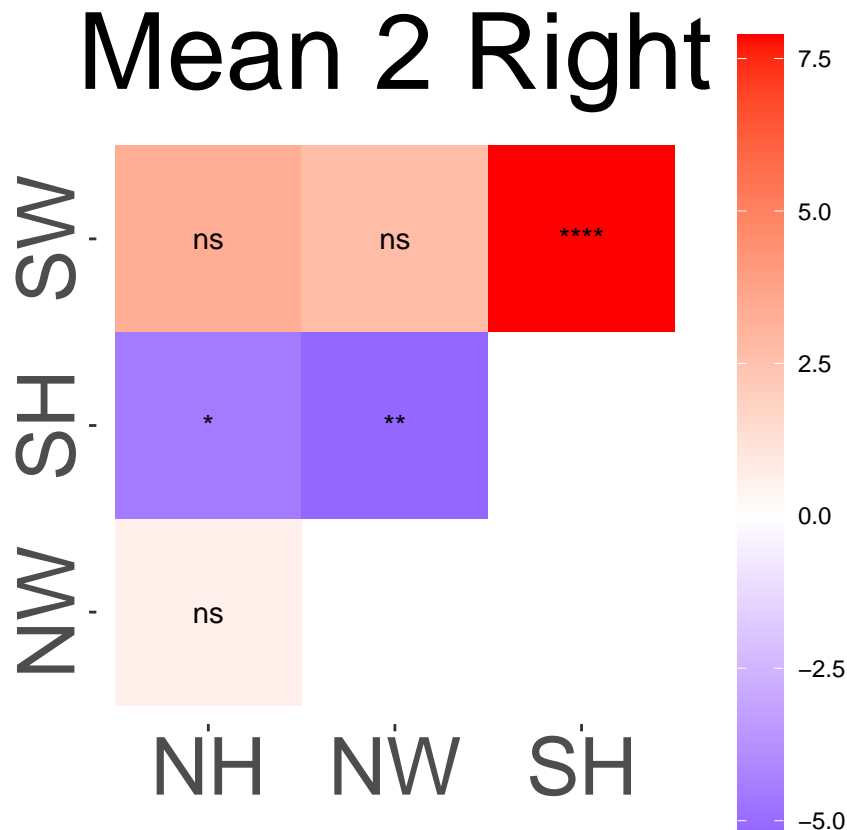


Plotting the heatmap of the right hemisphere data.

```
#Heatmap for Mean 2 right hemisphere
mid <- 0

ggplot(data = pwc_right_mean2, aes(x=group1, y=group2, fill=estimate))+
  geom_tile()+
  theme(plot.title = element_text(size = 40, face = "plain", hjust = 0.5),
        panel.background = element_blank(),
        axis.text.y = element_text(size = 30, face = "plain", angle = 90,
                                     hjust = 0.5),
        axis.text.x = element_text(size = 30, face = "plain"),
        legend.title = element_blank(),
        legend.key.height = unit(60, units = "pt"))+
  ggtitle("Mean 2 Right")+
  scale_fill_gradient2(low = "blue", high = "red", midpoint = mid,
                      mid = "white")+

  xlab("")+
  ylab("")+
  coord_equal()+
  geom_text(aes(label=p.adj.signif))
```



Step Seven: T-test of the left hemisphere data without FDR p-value adjustment.

```
#Mean 2 left pairwise t-test
pwc_left_mean2 <- pnn_data_left %>%
  pairwise_t_test(
    Mean.2 ~ Condition, pool.sd = FALSE,
    p.adjust.method = "none",
    detailed = TRUE
  )

pwc_left_mean2
```

```
## # A tibble: 6 x 17
##   estimate estimate1 estimate2 .y. group1 group2  n1  n2 statistic      p
## *   <dbl>      <dbl>      <dbl> <chr> <chr> <chr> <int> <int>    <dbl> <dbl>
## 1    0.881      69.6      68.7 Mean~ NH    NW     55   52     0.483 6.3 e-1
## 2   -6.04      69.6      75.7 Mean~ NH    SH     55   56    -2.88 5 e-3
## 3    4.00      69.6      65.6 Mean~ NH    SW     55   59     2.19 3.1 e-2
## 4   -6.92      68.7      75.7 Mean~ NW    SH     52   56    -4.06 9.9 e-5
## 5    3.12      68.7      65.6 Mean~ NW    SW     52   59     2.31 2.3 e-2
## 6   10.0      75.7      65.6 Mean~ SH    SW     56   59     5.89 5.37e-8
## # ... with 7 more variables: df <dbl>, conf.low <dbl>, conf.high <dbl>,
## #   method <chr>, alternative <chr>, p.adj <dbl>, p.adj.signif <chr>
```

## T-test of the right hemisphere data without FDR p-value adjustment.

```
#Mean 2 right pairwise t-test
pwc_right_mean2 <- pnn_data_right %>%
  pairwise_t_test(
    Mean.2 ~ Condition, pool.sd = FALSE,
    p.adjust.method = "none",
    detailed = TRUE
  )

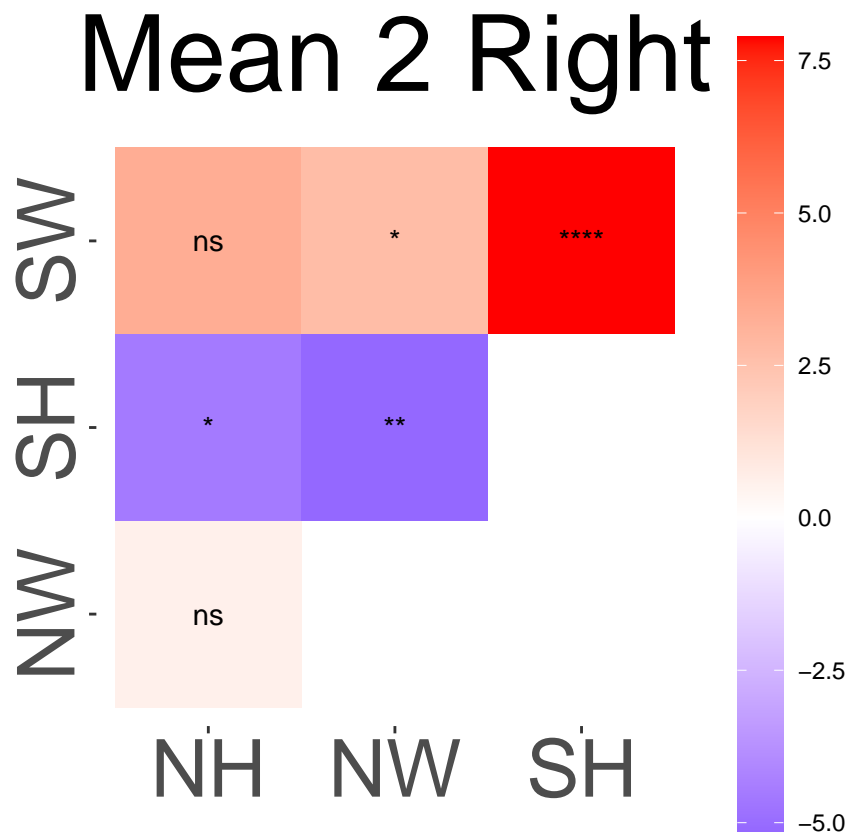
pwc_right_mean2

## # A tibble: 6 x 17
##   estimate estimate1 estimate2 .y. group1 group2  n1    n2 statistic      p
## *   <dbl>      <dbl>      <dbl> <chr> <chr>  <chr> <int> <int>    <dbl> <dbl>
## 1    0.612      69.8      69.2 Mean~ NH    NW     51    57     0.346 7.3 e-1
## 2   -4.54      69.8      74.3 Mean~ NH    SH     51    62    -2.36 2 e-2
## 3    3.34      69.8      66.5 Mean~ NH    SW     51    57     1.93 5.7 e-2
## 4   -5.15      69.2      74.3 Mean~ NW    SH     57    62    -3.36 1 e-3
## 5    2.72      69.2      66.5 Mean~ NW    SW     57    57     2.13 3.5 e-2
## 6    7.88      74.3      66.5 Mean~ SH    SW     62    57     5.29 6.32e-7
## # ... with 7 more variables: df <dbl>, conf.low <dbl>, conf.high <dbl>,
## #   method <chr>, alternative <chr>, p.adj <dbl>, p.adj.signif <chr>
```

## Step Eight: Plotting the heatmap of the non-adjusted right hemisphere data.

```
#Heatmap for Mean 2 right un-adjusted hemisphere
mid <- 0

ggplot(data = pwc_right_mean2, aes(x=group1, y=group2, fill=estimate))+
  geom_tile()+
  theme(plot.title = element_text(size = 40, face = "plain", hjust = 0.5),
        panel.background = element_blank(),
        axis.text.y = element_text(size = 30, face = "plain", angle = 90,
                                     hjust = 0.5),
        axis.text.x = element_text(size = 30, face = "plain"),
        legend.title = element_blank(),
        legend.key.height = unit(60, units = "pt"))+
  ggtitle("Mean 2 Right")+
  scale_fill_gradient2(low = "blue", high = "red", midpoint = mid,
                      mid = "white")+
  xlab("")+
  ylab("")+
  coord_equal()+
  geom_text(aes(label=p.adj.signif))
```

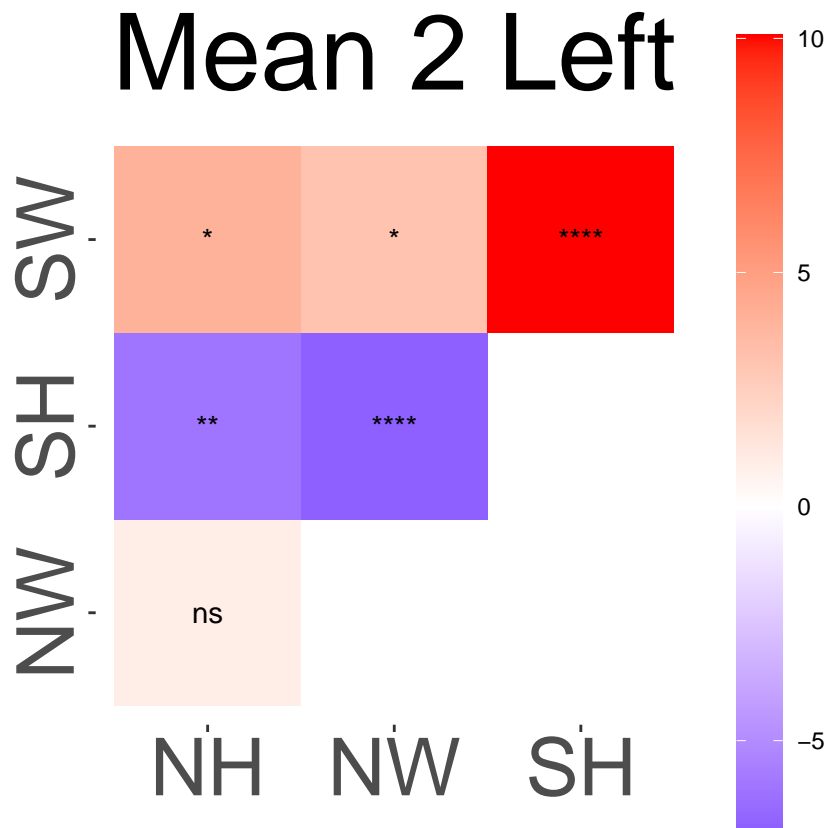


Plotting the heatmap of the non-adjusted left hemisphere data.

```
#Heatmap for Mean 2 left un-adjusted hemisphere
mid <- 0

ggplot(data = pwc_left_mean2, aes(x=group1, y=group2, fill=estimate))+
  geom_tile()+
  theme(plot.title = element_text(size = 40, face = "plain", hjust = 0.5),
        panel.background = element_blank(),
        axis.text.y = element_text(size = 30, face = "plain", angle = 90,
                                     hjust = 0.5),
        axis.text.x = element_text(size = 30, face = "plain"),
        legend.title = element_blank(),
        legend.key.height = unit(60, units = "pt"))+
  ggtitle("Mean 2 Left")+
  scale_fill_gradient2(low = "blue", high = "red", midpoint = mid,
                      mid = "white")+

  xlab("")+
  ylab("")+
  coord_equal()+
  geom_text(aes(label=p.adj.signif))
```



Step Nine: Due to differences between my results and previous results I have attempted to determine if my results make sense. I manually verify a couple of values that have flipped sign from previous results and still arrive at the calculated estimate of the t-test.

```
#Checking values that have opposite sign of previous results manually
pnn_data_right_nh <- filter(pnn_data_right, Condition=="NH")
mean(pnn_data_right_nh$Mean.2)

## [1] 69.7864

pnn_data_right_sh <- filter(pnn_data_right, Condition=="SH")
mean(pnn_data_right_sh$Mean.2)

## [1] 74.32697

diff_nh_sh <- mean(pnn_data_right_nh$Mean.2) - mean(pnn_data_right_sh$Mean.2)
diff_nh_sh

## [1] -4.540574

pnn_data_right_nw <- filter(pnn_data_right, Condition=="NW")
mean(pnn_data_right_nw$Mean.2)

## [1] 69.17413

diff_nw_sh <- mean(pnn_data_right_nw$Mean.2) - mean(pnn_data_right_sh$Mean.2)
diff_nw_sh
```



```
## [1] -5.152847
```

Step Ten: While I am confident about how I arrive at my result it is still not identical to previous work. When looking at previous code it looks like it calculates the difference in means by subtracting the first vector of estimates from the second vector of estimates (mean2 - mean1). By comparison, my code takes the first estimate minus the second estimate (mean1 - mean2). This could explain the difference in sign that I see. Checking it here. Once the t-test is done, I simply update the estimate by subtracting the second estimate vector from the first estimate vector and see if that gets me the identical results of previous code when it comes to the heatmap.

```
#Mean 2 right pairwise t-test
pwc_right_mean2 <- pnn_data_right %>%
  pairwise_t_test(
    Mean.2 ~ Condition, pool.sd = FALSE,
    p.adjust.method = "none",
    detailed = TRUE
  )

pwc_right_mean2

## # A tibble: 6 x 17
##   estimate estimate1 estimate2 .y. group1 group2 n1 n2 statistic p
## *   <dbl>      <dbl>      <dbl> <chr> <chr> <chr> <int> <int>    <dbl> <dbl>
## 1    0.612      69.8      69.2 Mean~ NH    NW     51    57     0.346 7.3 e-1
## 2   -4.54      69.8      74.3 Mean~ NH    SH     51    62    -2.36 2 e-2
## 3    3.34      69.8      66.5 Mean~ NH    SW     51    57     1.93 5.7 e-2
## 4   -5.15      69.2      74.3 Mean~ NW    SH     57    62    -3.36 1 e-3
## 5    2.72      69.2      66.5 Mean~ NW    SW     57    57     2.13 3.5 e-2
## 6    7.88      74.3      66.5 Mean~ SH    SW     62    57     5.29 6.32e-7
## # ... with 7 more variables: df <dbl>, conf.low <dbl>, conf.high <dbl>,
## #   method <chr>, alternative <chr>, p.adj <dbl>, p.adj.signif <chr>

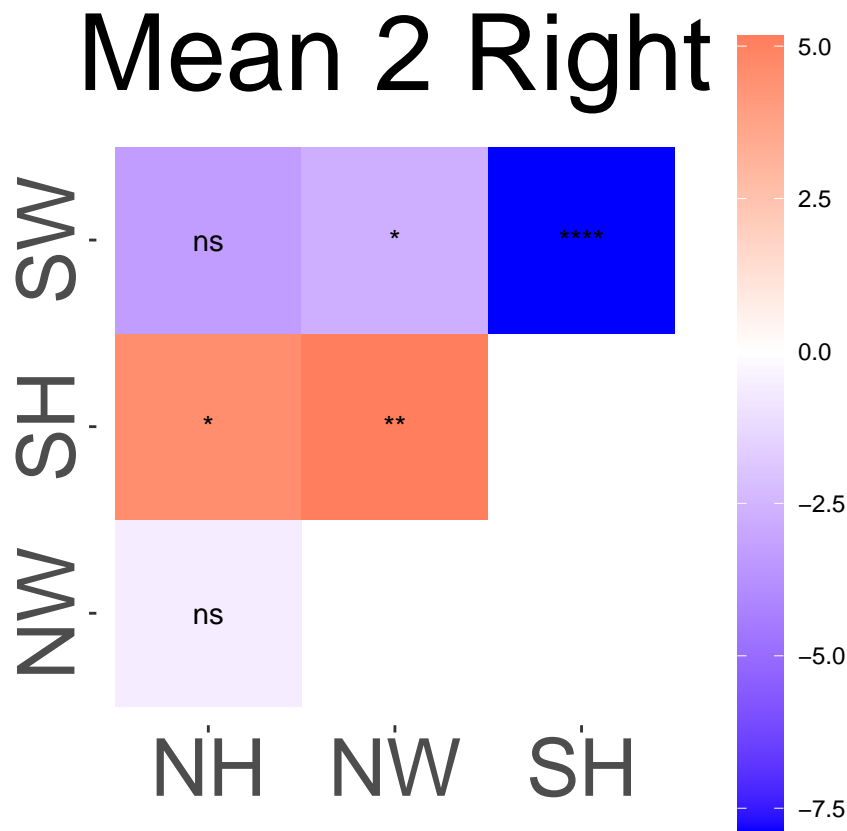
pwc_right_mean2$estimate <- pwc_right_mean2$estimate2 - pwc_right_mean2$estimate1
pwc_right_mean2$estimate

## [1] -0.612273  4.540574 -3.335321  5.152847 -2.723048 -7.875894
```

```
#Heatmap for Mean 2 right un-adjusted hemisphere
mid <- 0

ggplot(data = pwc_right_mean2, aes(x=group1, y=group2, fill=estimate))+
  geom_tile()+
  theme(plot.title = element_text(size = 40, face = "plain", hjust = 0.5),
        panel.background = element_blank(),
        axis.text.y = element_text(size = 30, face = "plain", angle = 90,
                                     hjust = 0.5),
        axis.text.x = element_text(size = 30, face = "plain"),
        legend.title = element_blank(),
        legend.key.height = unit(60, units = "pt"))+
  ggtitle("Mean 2 Right")+
  scale_fill_gradient2(low = "blue", high = "red", midpoint = mid,
                      mid = "white")+
```

```
xlab("")+
ylab("")+
coord_equal()+
geom_text(aes(label=p.adj.signif))
```



Step Eleven: This fixes two of the results but still causes problems for others. Will now try and manually specify the exact combinations of tests that should be performed to get exactly the same output

```
#Mean 2 right pairwise t-test
pwc_right_mean2 <- pnn_data_right %>%
  pairwise_t_test(
    Mean.2 ~ Condition, pool.sd = FALSE,
    p.adjust.method = "none",
    detailed = TRUE, comparisons = list(c("NW", "NH"),
                                         c("SW", "NH"), c("NW", "SW"),
                                         c("NH", "SW"), c("NW", "SH"),
                                         c("NH", "SH"), c("SW", "SH"))
  )

pwc_right_mean2

## # A tibble: 7 x 17
##   estimate estimate1 estimate2 .y. group1 group2  n1    n2 statistic      p
## *   <dbl>     <dbl>     <dbl> <chr> <chr>  <chr> <int> <int>     <dbl> <dbl>
```

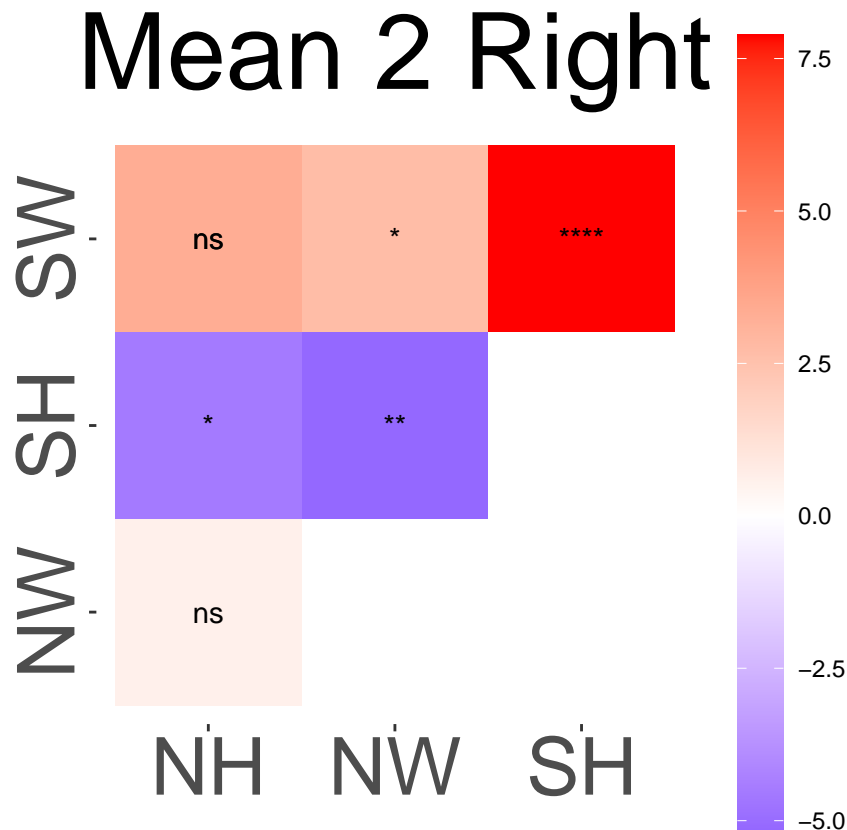
```
## 1    0.612      69.8      69.2 Mean~ NH      NW      51      57      0.346 7.3 e-1
## 2    3.34      69.8      66.5 Mean~ NH      SW      51      57      1.93  5.7 e-2
## 3    2.72      69.2      66.5 Mean~ NW      SW      57      57      2.13  3.5 e-2
## 4    3.34      69.8      66.5 Mean~ NH      SW      51      57      1.93  5.7 e-2
## 5   -5.15      69.2      74.3 Mean~ NW      SH      57      62     -3.36  1   e-3
## 6   -4.54      69.8      74.3 Mean~ NH      SH      51      62     -2.36  2   e-2
## 7    7.88      74.3      66.5 Mean~ SH      SW      62      57      5.29  6.32e-7
## # ... with 7 more variables: df <dbl>, conf.low <dbl>, conf.high <dbl>,
## #   method <chr>, alternative <chr>, p.adj <dbl>, p.adj.signif <chr>
```

## Step Twelve: Plotting the heatmap of specifically defined comparisons

```
#Heatmap for Mean 2 right un-adjusted hemisphere
mid <- 0

ggplot(data = pwc_right_mean2, aes(x=group1, y=group2, fill=estimate))+
  geom_tile()+
  theme(plot.title = element_text(size = 40, face = "plain", hjust = 0.5),
        panel.background = element_blank(),
        axis.text.y = element_text(size = 30, face = "plain", angle = 90,
                                    hjust = 0.5),
        axis.text.x = element_text(size = 30, face = "plain"),
        legend.title = element_blank(),
        legend.key.height = unit(60, units = "pt"))+
  ggtitle("Mean 2 Right")+
  scale_fill_gradient2(low = "blue", high = "red", midpoint = mid,
                      mid = "white")+

  xlab("")+
  ylab("")+
  coord_equal()+
  geom_text(aes(label=p.adj.signif))
```



Conclusion: I stand by my results with the differences in sign. I have gone back and manually verified the calculation that is arrived at by the pairwise t-test and I am confident that they are accurate. Based on my investigation it appears that the differences in sign are because of how the pairwise t-test is performed between this analysis and previous analysis. In the previous analysis it is the difference of mean2-mean1 while in this one it is mean1-mean2 and therefore leads to a difference in sign because of that. Even when manually specifying the combinations my results stay consistent. Given all of this I feel comfortable standing behind my results. Using the FDR p-value adjustment will be needed considering we have nearly a 1/5 chance of having at least one false positive in our results.