Data analysis

1. Read Data

```
library(readxl)
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(ggplot2)
caa <- read_excel("~/Desktop/MA675/Consulting/caa_all_radii_40um_donut_13Oct2025.xlsx")
head(caa)
```

```
# A tibble: 6 x 5
  Groups       Region subID distance OpticalProperty
  <chr>        <chr>  <dbl>    <dbl>           <dbl>
1 experimental front      1       40            11.4
2 experimental front      1       40            12.9
3 experimental front      1       40            13.1
4 experimental front      1       40            12.9
5 experimental front      1       40            12.6
6 experimental front      1       40            13.2
```

```
summary(caa)
```

```
    Groups             Region               subID          distance     
 Length:330787      Length:330787      Min.   :1.000   Min.   : 40.0  
 Class :character   Class :character   1st Qu.:3.000   1st Qu.: 40.0  
 Mode  :character   Mode  :character   Median :4.000   Median :120.0  
                                       Mean   :4.671   Mean   :144.4  
                                       3rd Qu.:7.000   3rd Qu.:200.0  
                                       Max.   :9.000   Max.   :480.0  
 OpticalProperty  
 Min.   : 0.5275  
 1st Qu.:10.8702  
 Median :12.1511  
 Mean   :11.8954  
 3rd Qu.:13.0886  
 Max.   :82.9646  
```

```
caa %>% count(Groups, Region)
```

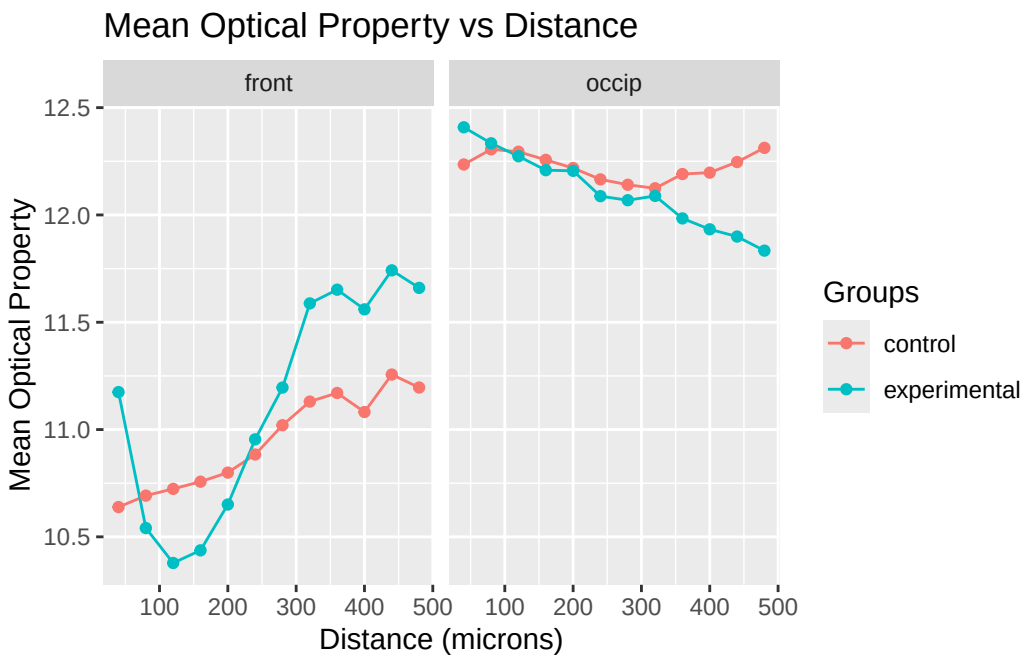
```
# A tibble: 4 x 3
  Groups       Region      n
  <chr>        <chr>   <int>
1 control      front   75264
2 control      occip  245446
3 experimental front    4731
4 experimental occip    5346
```

First, I checked the data structure. There are measurements for five subjects across frontal and occipital regions, and for both EPVS and normal vessels. The data includes optical values over distances from the vessel boundary.

2

```
avg_caa <- caa %>%
  group_by(Groups, Region, distance) %>%
  summarise(mean_opt = mean(OpticalProperty, na.rm = TRUE),
            sd_opt = sd(OpticalProperty), .groups="drop")
```

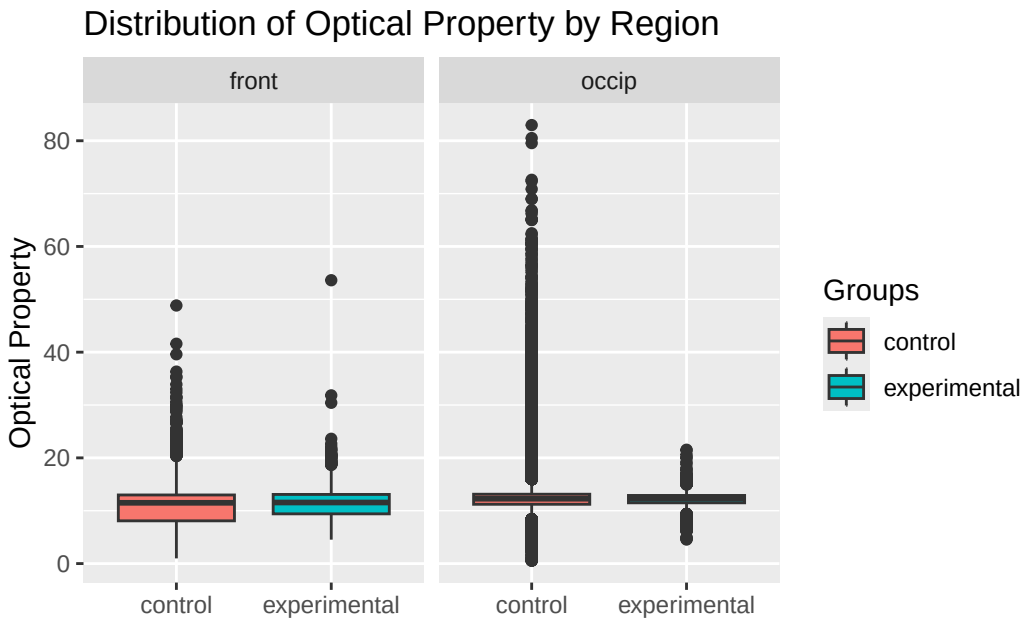
```
ggplot(avg_caa, aes(x = distance, y = mean_opt, color = Groups)) +
  geom_line() +
  geom_point() +
  facet_wrap(~Region) +
  labs(title="Mean Optical Property vs Distance",
       y="Mean Optical Property",
       x="Distance (microns)")
```



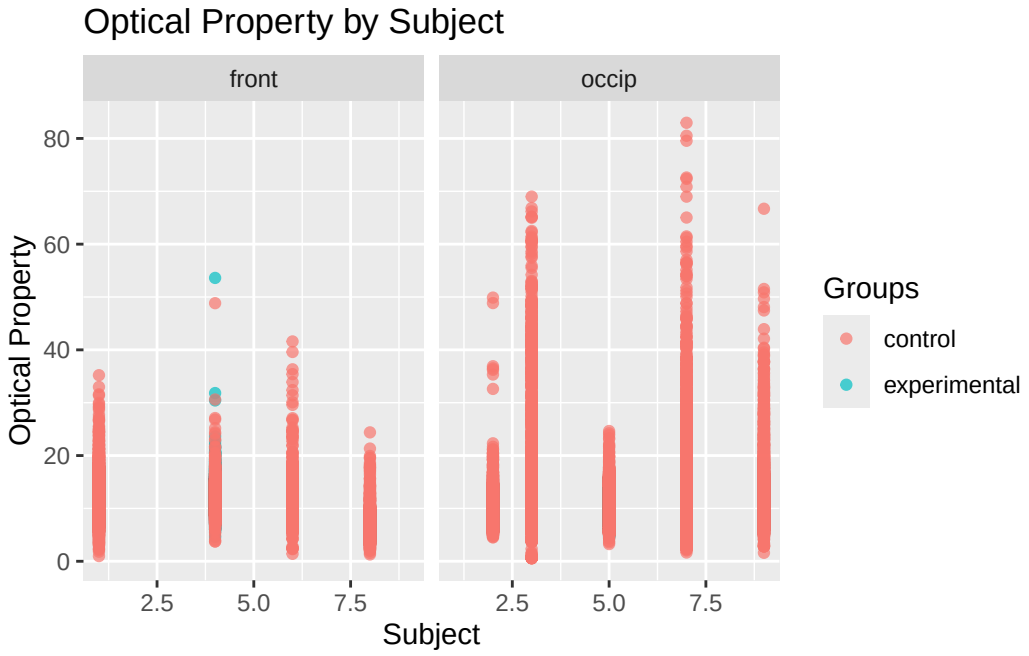
I plotted the mean optical values across distance for EPVS vs normal vessels. We see a general difference between EPVS and healthy vessels, especially in the occipital region. The EPVS curve shows lower values near the boundary, consistent with expected myelin loss.

3

```
ggplot(caa, aes(x=Groups, y=OpticalProperty, fill=Groups)) +
  geom_boxplot() +
  facet_wrap(~Region) +
  labs(title="Distribution of Optical Property by Region",
       y="Optical Property", x="")
```




```
ggplot(caa, aes(x=subID, y=OpticalProperty, color=Groups)) +
  geom_point(alpha=0.7) +
  facet_wrap(~Region) +
  labs(title="Optical Property by Subject",
       y="Optical Property", x="Subject")
```



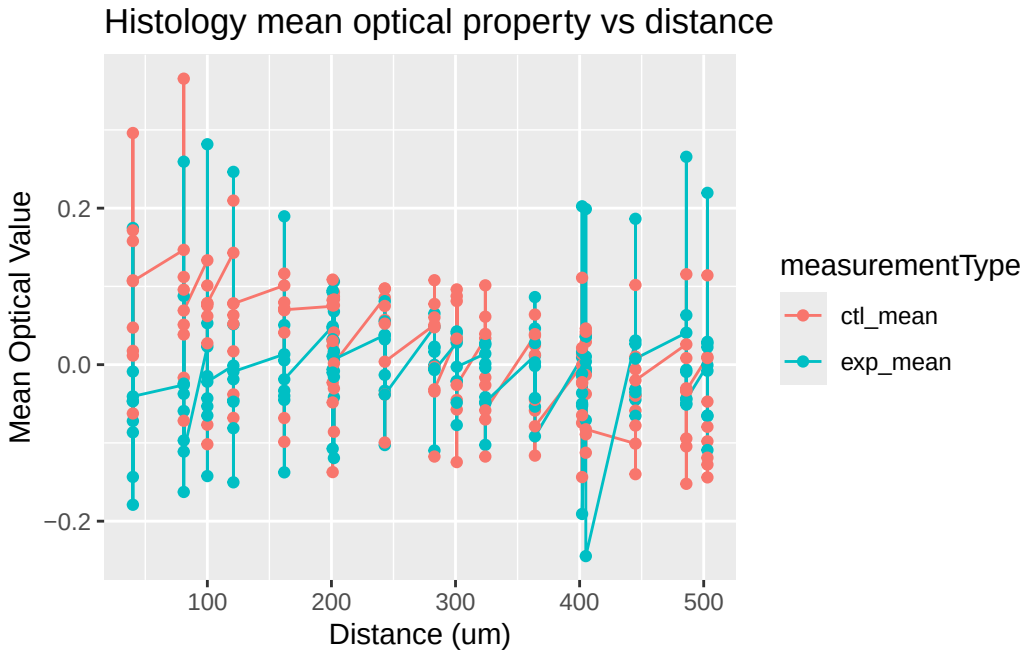
Read Data for histology

```
hist_mean <- read_excel("~/Desktop/MA675/Consulting/histology_mean_ring_measurements_15-Oct-2
head(hist_mean)
```

```
# A tibble: 6 x 5
  baseName radiusSize radiusValue measurementType measurement
  <chr>    <chr>      <chr>       <chr>                  <dbl>
1 CAA_17_7 rad40      rad40       exp_mean             -0.0865
2 CAA_17_7 rad40      rad40       ctl_mean              0.158
3 CAA_17_7 rad40      rad81       exp_mean             -0.0259
4 CAA_17_7 rad40      rad81       ctl_mean              0.147
5 CAA_17_7 rad40      rad121      exp_mean             -0.00115
6 CAA_17_7 rad40      rad121      ctl_mean              0.143
```

```
hist_mean$distance <- as.numeric(gsub("rad","", hist_mean$radiusValue))

ggplot(hist_mean, aes(x = distance, y = measurement, color = measurementType)) +
  geom_line() +
  geom_point() +
  ggtitle("Histology mean optical property vs distance") +
  xlab("Distance (um)") +
  ylab("Mean Optical Value")
```



To simplify the visualization, I also plotted the average curve for each group across distances. The EPVS curve stays lower than the control curve at most distances, matching the OCT trend.