PLS_C_analysis

Quin

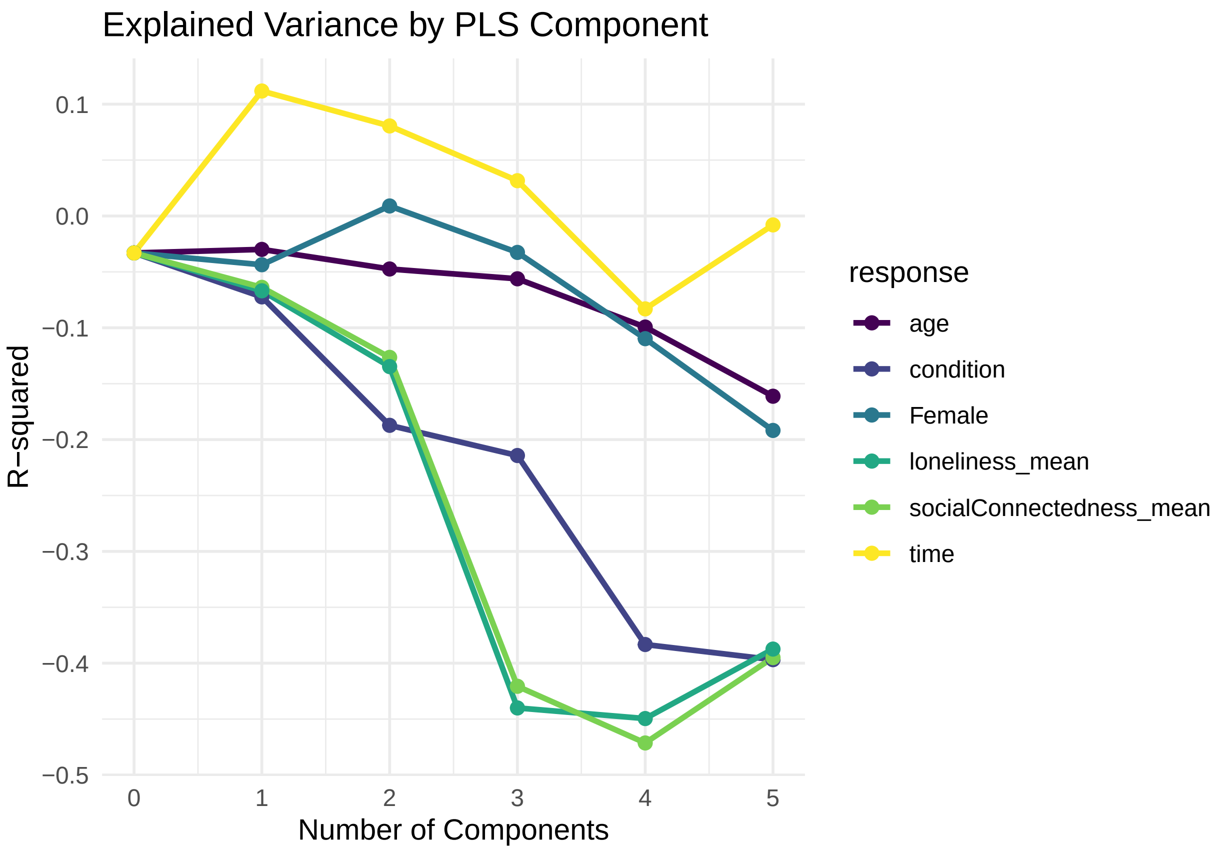
2025-05-29

```r
# Standardize the data
connectivity_features_scaled <- scale(connectivity_features_selected)
behavioral_data_scaled <- scale(behavioral_data)

# Run PLS-C analysis
pls_result <- plsr(behavioral_data_scaled ~ connectivity_features_scaled,
                   ncomp = min(5, ncol(behavioral_data_scaled)),
                   validation = "CV",
                   method = "oscorespls")
summary(pls_result)
```

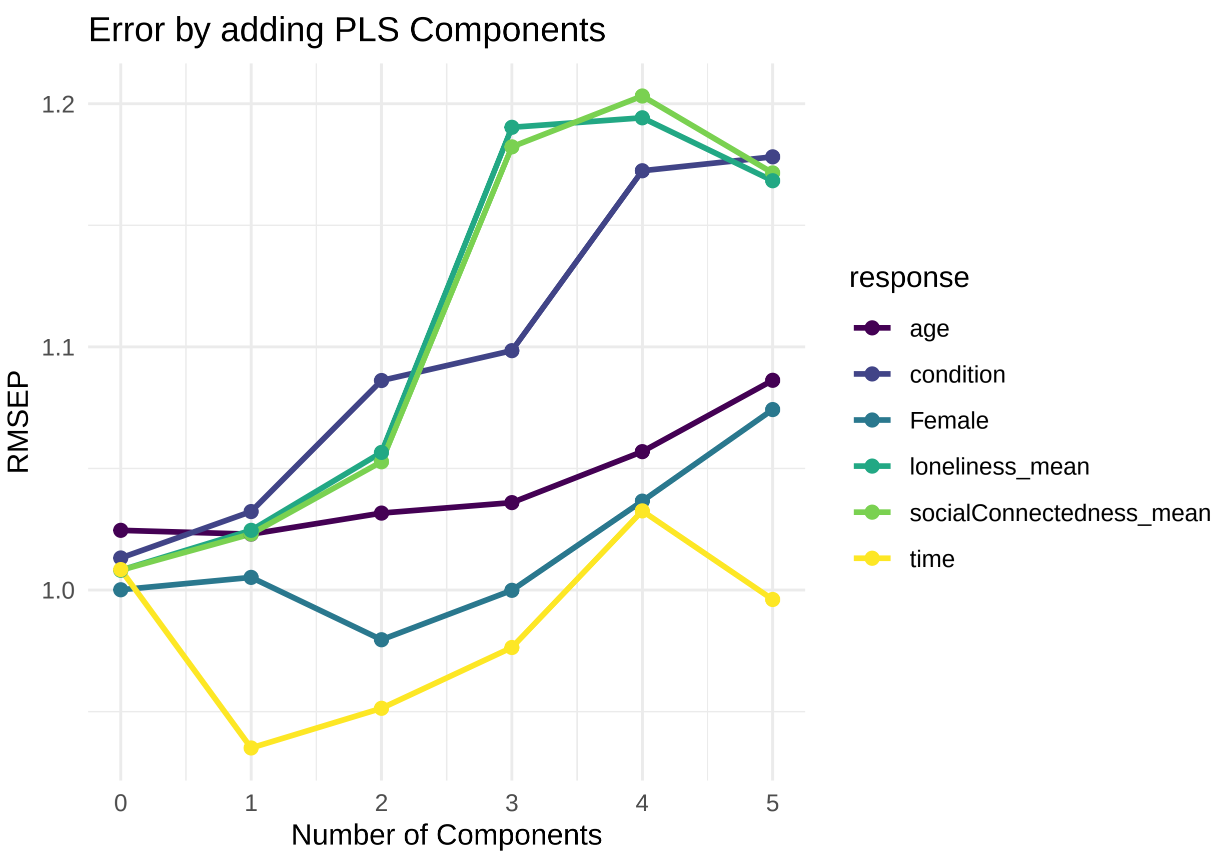
```
## Data:    X dimension: 62 2346
##  Y dimension: 62 6
## Fit method: oscorespls
## Number of components considered: 5
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##
## Response: condition
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
## CV           1.013    1.032    1.086    1.098    1.172    1.178
## adjCV        1.013    1.031    1.081    1.097    1.153    1.152
##
## Response: age
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
## CV           1.025    1.023    1.032    1.036    1.057    1.086
## adjCV        1.025    1.021    1.030    1.033    1.056    1.118
##
## Response: Female
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
## CV               1    1.005   0.9796   0.9999    1.037    1.074
## adjCV            1    1.004   0.9772   0.9908    1.030    1.049
##
## Response: socialConnectedness_mean
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
## CV           1.008    1.023    1.053    1.182    1.203    1.172
## adjCV        1.008    1.022    1.049    1.160    1.162    1.133
##
## Response: loneliness_mean
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
## CV           1.008    1.025    1.057    1.190    1.194    1.168
## adjCV        1.008    1.023    1.054    1.168    1.159    1.149
```

```
##
## Response: time
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
## CV           1.008   0.9351   0.9514   0.9764    1.033   0.9961
## adjCV        1.008   0.9333   0.9480   0.9777    1.034   0.9674
##
## TRAINING: % variance explained
##                          1 comps  2 comps  3 comps  4 comps  5 comps
## X                       40.680901   53.843   58.048   61.252   64.569
## condition                0.002043    3.540    4.590   25.022   33.321
## age                      2.735885    2.821    4.067    4.484    4.725
## Female                   0.396046   11.681   17.934   20.304   32.360
## socialConnectedness_mean  0.090943    9.751   32.650   44.081   46.718
## loneliness_mean          0.543853    3.292   27.542   35.051   35.153
## time                    18.258197   24.374   24.980   31.834   48.982
```

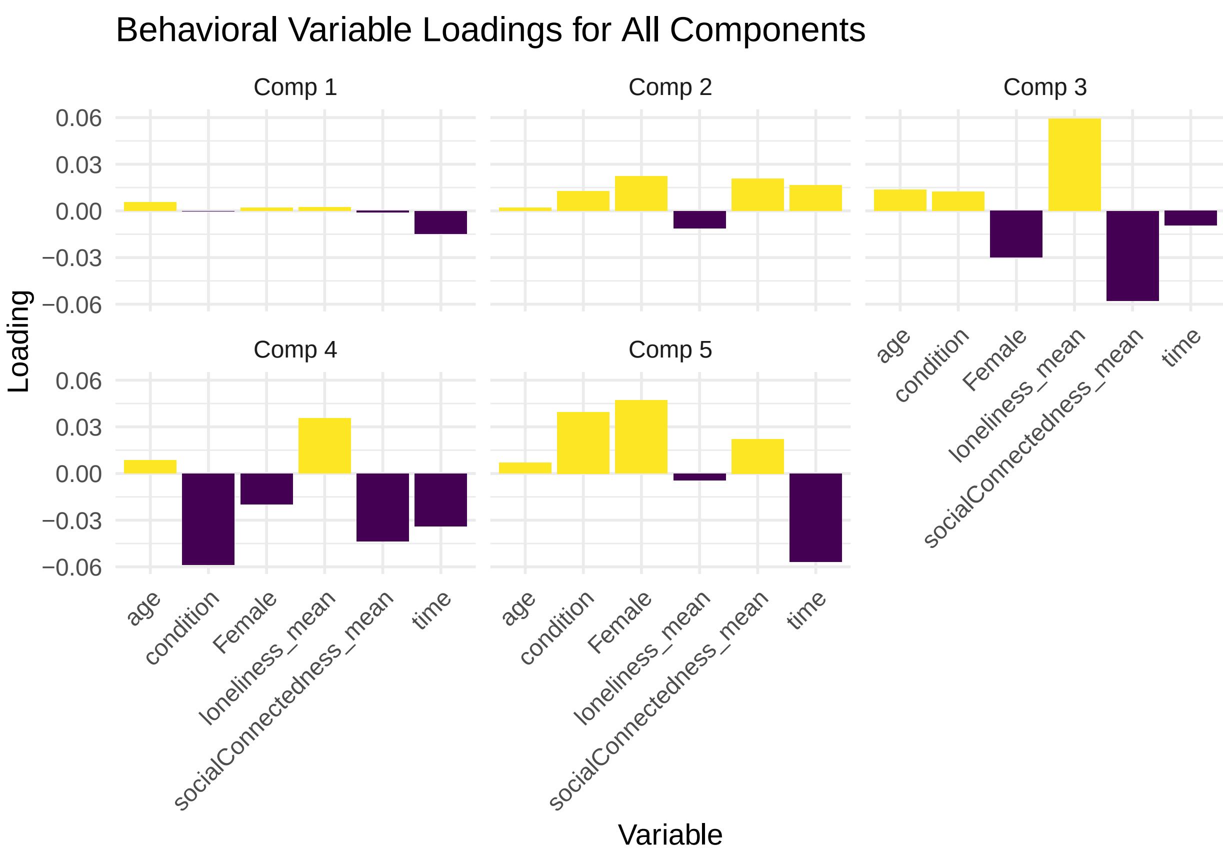
```r
# Plot explained variance by number of components
explained_var <- as.data.frame(R2(pls_result)) %>%
  rownames_to_column("Component")

ggplot(explained_var, aes(x = comps, y = value, color = response, group = response))+
  geom_line(size=1) +
  geom_point(size=2) +
  theme_minimal() +
  labs(title = "Explained Variance by PLS Component",
       x = "Number of Components",
       y = "R-squared")+
  scale_color_viridis_d()
```



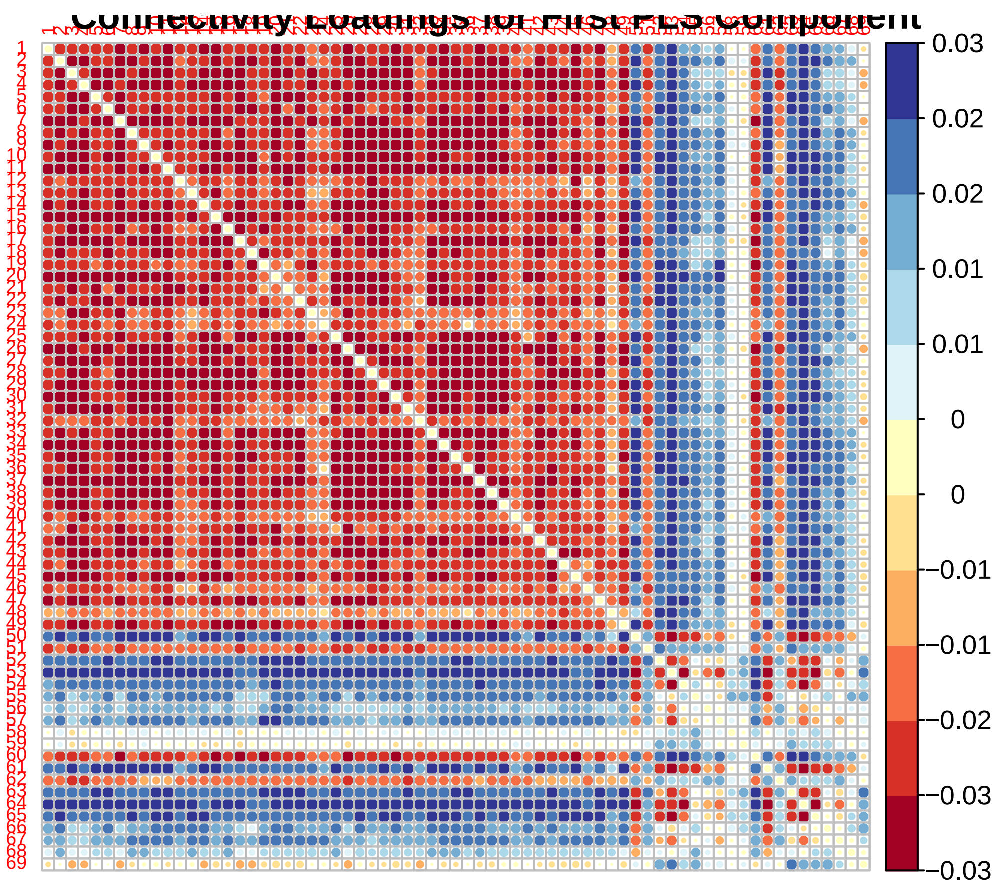

```r
rmse <- as.data.frame(RMSEP(pls_result)) %>%
  rownames_to_column("Component")
rmse$response = rep(rep(c("condition","age","Female","socialConnectedness_mean","loneliness_mean","time

rmse%>%filter(estimate == "CV")%>%ggplot(aes(x = comps, y = value, color = response, group = response))
  geom_line(size=1) +
  geom_point(size=2) +
  theme_minimal() +
  labs(title = "Error by adding PLS Components",
       x = "Number of Components",
       y = "RMSEP")+
  scale_color_viridis_d()
```



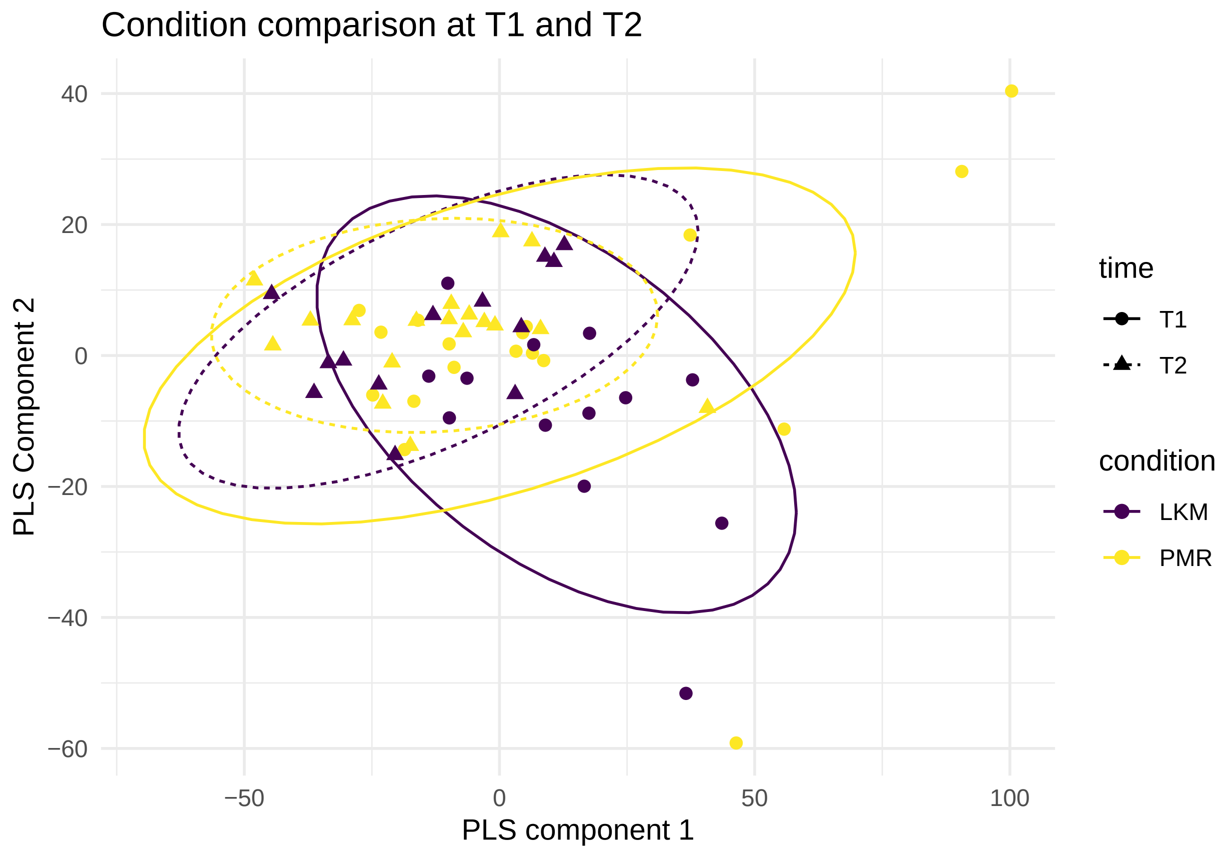

```r
# plot variable loadings for behavioral variables
loadings_behavior <- Yloadings(pls_result)
loadings_behavior_df <- as.data.frame(loadings_behavior[,1:5]) %>%
  rownames_to_column("Variable") %>%
  pivot_longer(-Variable, names_to = "Component", values_to = "Loading")

ggplot(loadings_behavior_df, aes(x = Variable, y = Loading,fill = (Loading>0))) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Behavioral Variable Loadings for All Components",
       x = "Variable", y = "Loading") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+facet_wrap(~Component,nrow = 2)+scale_fill_vi
```




```r
# connectivity loadings for the first component
# These represent the brain connections most associated with the behavioral variables
loadings_connectivity <- pls_result$loadings[,1]

# Convert loadings back to matrix form for visualization
loading_matrix <- matrix(0, nrow = 69, ncol = 69)
loading_matrix[upper.tri(loading_matrix)] <- loadings_connectivity
loading_matrix <- loading_matrix + t(loading_matrix)

# Plot the loading matrix
corrplot::corrplot(loading_matrix, # method = "circle",
         # type = "p",
         col=brewer.pal(n=11, name="RdYlBu"),
         title = "Connectivity Loadings for First PLS Component",
         is.corr = FALSE , tl.cex = 0.6
         )
```




```r
# Extract PLS scores for the first component
pls_scores <- scores(pls_result)[,1:2]

# Create dataframe for plotting
scores_df <- data.frame(behavioral_data_scaled[-c(32,64),],pls_scores)%>%mutate(condition = ifelse(condi

# Plot relationship
ggplot(scores_df, aes(x = Comp.1, y = Comp.2,color = condition, shape = time,group = interaction(time, c
  stat_ellipse(aes(linetype = time))+
  geom_point(size = 2) +
  theme_minimal() +
  labs(title = "Condition comparison at T1 and T2",
       x = "PLS component 1",
       y = "PLS Component 2")+
  scale_color_viridis_d()
```




```r
# Calculate correlation
cor_test <- cor.test(scores_df$loneliness_mean, scores_df$Comp.2)
print(paste("Correlation between loneliness and PLS Component 1:",
            round(cor_test$estimate, 3),
            "p-value:", round(cor_test$p.value, 4)))
```

```
## [1] "Correlation between loneliness and PLS Component 1: -0.166 p-value: 0.1979"
```