Distribution of Factors Across Batches

# Set up knitr  
knitr::opts\_chunk$set(echo = TRUE)  
  
# Load the libraries  
library(dplyr)

library(ggplot2)

library(knitr)

df <- read.table("Z:/ResearchHome/ClusterHome/aneupane/data/Yadav\_serum/v12\_output/CAB\_randomized\_batch1\_1200\_samples/Re\_ SJLIFE samples for proteomics\_metabolomics experiments/Output\_for\_Experiment\_Design\_\_v0.0.1.txt", header = T, sep = "\t")  
df <- df[!grepl("IR", df$sampleID),]  
  
df$batchID <- factor(df$batchID, levels = unique(c(df$batchID)))  
  
# 1. Check distribution of factor1 across batches  
factor1\_batch\_table <- table(df$factor1, df$batchID)  
# print(factor1\_batch\_table)  
knitr::kable(t(factor1\_batch\_table), caption = "Distribution of Factor 1 Across Batches", format = "markdown")

Distribution of Factor 1 Across Batches

|  | 100\_community\_controls | 171\_CMP\_cases | 200\_Hodgkin\_lymphoma | 729\_non\_Hodgkin |
| --- | --- | --- | --- | --- |
| batch1 | 2 | 2 | 2 | 9 |
| batch2 | 2 | 2 | 2 | 9 |
| batch3 | 2 | 2 | 2 | 9 |
| batch4 | 2 | 2 | 2 | 9 |
| batch5 | 2 | 2 | 2 | 9 |
| batch6 | 2 | 2 | 2 | 9 |
| batch7 | 2 | 2 | 2 | 9 |
| batch8 | 2 | 2 | 2 | 9 |
| batch9 | 2 | 2 | 2 | 9 |
| batch10 | 2 | 2 | 2 | 9 |
| batch11 | 2 | 2 | 2 | 9 |
| batch12 | 2 | 2 | 2 | 9 |
| batch13 | 2 | 2 | 2 | 9 |
| batch14 | 2 | 2 | 2 | 9 |
| batch15 | 2 | 2 | 2 | 9 |
| batch16 | 2 | 2 | 2 | 9 |
| batch17 | 2 | 2 | 2 | 9 |
| batch18 | 2 | 2 | 2 | 9 |
| batch19 | 2 | 2 | 2 | 9 |
| batch20 | 2 | 2 | 2 | 9 |
| batch21 | 1 | 3 | 2 | 9 |
| batch22 | 1 | 3 | 2 | 9 |
| batch23 | 1 | 3 | 2 | 9 |
| batch24 | 1 | 3 | 2 | 9 |
| batch25 | 1 | 3 | 2 | 9 |
| batch26 | 1 | 3 | 2 | 9 |
| batch27 | 1 | 3 | 2 | 9 |
| batch28 | 1 | 3 | 2 | 9 |
| batch29 | 1 | 3 | 2 | 9 |
| batch30 | 1 | 3 | 2 | 9 |
| batch31 | 1 | 3 | 2 | 9 |
| batch32 | 1 | 2 | 3 | 9 |
| batch33 | 1 | 2 | 3 | 9 |
| batch34 | 1 | 2 | 3 | 9 |
| batch35 | 1 | 2 | 3 | 9 |
| batch36 | 1 | 2 | 3 | 9 |
| batch37 | 1 | 2 | 3 | 9 |
| batch38 | 1 | 2 | 3 | 9 |
| batch39 | 1 | 2 | 3 | 9 |
| batch40 | 1 | 2 | 3 | 9 |
| batch41 | 1 | 2 | 3 | 9 |
| batch42 | 1 | 2 | 3 | 9 |
| batch43 | 1 | 2 | 3 | 9 |
| batch44 | 1 | 2 | 3 | 9 |
| batch45 | 1 | 2 | 3 | 9 |
| batch46 | 1 | 2 | 3 | 9 |
| batch47 | 1 | 2 | 3 | 9 |
| batch48 | 1 | 2 | 3 | 9 |
| batch49 | 1 | 2 | 3 | 9 |
| batch50 | 1 | 2 | 3 | 9 |
| batch51 | 1 | 2 | 3 | 9 |
| batch52 | 1 | 2 | 3 | 9 |
| batch53 | 1 | 2 | 3 | 9 |
| batch54 | 1 | 2 | 3 | 9 |
| batch55 | 1 | 2 | 3 | 9 |
| batch56 | 1 | 2 | 3 | 9 |
| batch57 | 1 | 2 | 3 | 9 |
| batch58 | 1 | 2 | 3 | 9 |
| batch59 | 1 | 2 | 3 | 9 |
| batch60 | 1 | 2 | 3 | 9 |
| batch61 | 1 | 2 | 3 | 9 |
| batch62 | 1 | 2 | 3 | 9 |
| batch63 | 1 | 2 | 3 | 9 |
| batch64 | 1 | 2 | 3 | 9 |
| batch65 | 1 | 2 | 3 | 9 |
| batch66 | 1 | 2 | 3 | 9 |
| batch67 | 1 | 2 | 3 | 9 |
| batch68 | 1 | 2 | 3 | 9 |
| batch69 | 1 | 2 | 3 | 9 |
| batch70 | 1 | 2 | 3 | 9 |
| batch71 | 1 | 2 | 3 | 9 |
| batch72 | 1 | 2 | 2 | 10 |
| batch73 | 1 | 2 | 2 | 10 |
| batch74 | 1 | 2 | 2 | 10 |
| batch75 | 1 | 2 | 2 | 10 |
| batch76 | 1 | 2 | 2 | 10 |
| batch77 | 1 | 2 | 2 | 10 |
| batch78 | 1 | 2 | 2 | 10 |
| batch79 | 1 | 2 | 2 | 10 |
| batch80 | 1 | 2 | 2 | 10 |

library(ggplot2)  
  
# Creating proportional table for each factor  
## Factor 1  
prop\_table <- prop.table(factor1\_batch\_table, margin = 2)  
# Convert to a data frame for ggplot2  
prop\_df <- as.data.frame(as.table(prop\_table))  
colnames(prop\_df) <- c("Factor1", "Batch", "Proportion")  
  
# Create the stacked bar plot  
f1 <- ggplot(prop\_df, aes(x = Batch, y = Proportion, fill = Factor1)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Proportion of Selection-group Across Batches", x = "Batch", y = "Proportion") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))  
f1

A graph of different colored lines

Description automatically generated with medium confidence

# 2. Check distribution of factor2 across batches  
factor2\_batch\_table <- table(df$factor2, df$batchID)  
# print(factor2\_batch\_table)  
knitr::kable(t(factor2\_batch\_table), caption = "Distribution of Factor 2 Across Batches", format = "markdown")

Distribution of Factor 2 Across Batches

|  | Female | Male |
| --- | --- | --- |
| batch1 | 9 | 6 |
| batch2 | 9 | 6 |
| batch3 | 9 | 6 |
| batch4 | 9 | 6 |
| batch5 | 9 | 6 |
| batch6 | 9 | 6 |
| batch7 | 9 | 6 |
| batch8 | 9 | 6 |
| batch9 | 9 | 6 |
| batch10 | 9 | 6 |
| batch11 | 9 | 6 |
| batch12 | 9 | 6 |
| batch13 | 9 | 6 |
| batch14 | 9 | 6 |
| batch15 | 9 | 6 |
| batch16 | 9 | 6 |
| batch17 | 9 | 6 |
| batch18 | 9 | 6 |
| batch19 | 9 | 6 |
| batch20 | 9 | 6 |
| batch21 | 9 | 6 |
| batch22 | 9 | 6 |
| batch23 | 9 | 6 |
| batch24 | 9 | 6 |
| batch25 | 9 | 6 |
| batch26 | 9 | 6 |
| batch27 | 9 | 6 |
| batch28 | 9 | 6 |
| batch29 | 9 | 6 |
| batch30 | 9 | 6 |
| batch31 | 9 | 6 |
| batch32 | 9 | 6 |
| batch33 | 9 | 6 |
| batch34 | 9 | 6 |
| batch35 | 9 | 6 |
| batch36 | 8 | 7 |
| batch37 | 8 | 7 |
| batch38 | 8 | 7 |
| batch39 | 8 | 7 |
| batch40 | 7 | 8 |
| batch41 | 7 | 8 |
| batch42 | 7 | 8 |
| batch43 | 7 | 8 |
| batch44 | 7 | 8 |
| batch45 | 7 | 8 |
| batch46 | 7 | 8 |
| batch47 | 7 | 8 |
| batch48 | 7 | 8 |
| batch49 | 7 | 8 |
| batch50 | 7 | 8 |
| batch51 | 7 | 8 |
| batch52 | 7 | 8 |
| batch53 | 7 | 8 |
| batch54 | 7 | 8 |
| batch55 | 7 | 8 |
| batch56 | 6 | 9 |
| batch57 | 6 | 9 |
| batch58 | 6 | 9 |
| batch59 | 6 | 9 |
| batch60 | 6 | 9 |
| batch61 | 6 | 9 |
| batch62 | 6 | 9 |
| batch63 | 6 | 9 |
| batch64 | 6 | 9 |
| batch65 | 6 | 9 |
| batch66 | 6 | 9 |
| batch67 | 5 | 10 |
| batch68 | 5 | 10 |
| batch69 | 5 | 10 |
| batch70 | 5 | 10 |
| batch71 | 5 | 10 |
| batch72 | 5 | 10 |
| batch73 | 5 | 10 |
| batch74 | 5 | 10 |
| batch75 | 5 | 10 |
| batch76 | 5 | 10 |
| batch77 | 5 | 10 |
| batch78 | 5 | 10 |
| batch79 | 5 | 10 |
| batch80 | 5 | 10 |

## Factor 2  
prop\_table <- prop.table(factor2\_batch\_table, margin = 2)  
# Convert to a data frame for ggplot2  
prop\_df <- as.data.frame(as.table(prop\_table))  
colnames(prop\_df) <- c("Factor2", "Batch", "Proportion")  
  
# Create the stacked bar plot  
f2 <- ggplot(prop\_df, aes(x = Batch, y = Proportion, fill = Factor2)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Proportion of Sex Across Batches", x = "Batch", y = "Proportion") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))  
f2

A graph of a number of people

Description automatically generated with medium confidence

# 3. Check distribution of factor3 across batches  
factor3\_batch\_table <- table(df$factor3, df$batchID)  
# print(factor3\_batch\_table)  
knitr::kable(t(factor3\_batch\_table), caption = "Distribution of Factor 3 Across Batches", format = "markdown")

Distribution of Factor 3 Across Batches

|  | [18,24] | (24,30] | (30,36] | (36,42] | (42,48] | (48,54] | (54,60] | (60,66] |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| batch1 | 2 | 5 | 5 | 2 | 1 | 0 | 0 | 0 |
| batch2 | 4 | 2 | 5 | 2 | 2 | 0 | 0 | 0 |
| batch3 | 4 | 2 | 4 | 4 | 0 | 1 | 0 | 0 |
| batch4 | 3 | 2 | 5 | 4 | 1 | 0 | 0 | 0 |
| batch5 | 5 | 2 | 2 | 0 | 3 | 3 | 0 | 0 |
| batch6 | 4 | 5 | 2 | 3 | 1 | 0 | 0 | 0 |
| batch7 | 2 | 5 | 5 | 2 | 0 | 1 | 0 | 0 |
| batch8 | 3 | 4 | 3 | 3 | 2 | 0 | 0 | 0 |
| batch9 | 4 | 3 | 4 | 1 | 1 | 2 | 0 | 0 |
| batch10 | 6 | 4 | 4 | 1 | 0 | 0 | 0 | 0 |
| batch11 | 4 | 4 | 4 | 2 | 1 | 0 | 0 | 0 |
| batch12 | 4 | 2 | 4 | 0 | 2 | 3 | 0 | 0 |
| batch13 | 3 | 3 | 5 | 2 | 2 | 0 | 0 | 0 |
| batch14 | 2 | 5 | 4 | 3 | 1 | 0 | 0 | 0 |
| batch15 | 6 | 0 | 4 | 2 | 1 | 2 | 0 | 0 |
| batch16 | 4 | 4 | 3 | 1 | 2 | 0 | 1 | 0 |
| batch17 | 3 | 3 | 5 | 2 | 1 | 1 | 0 | 0 |
| batch18 | 4 | 5 | 1 | 3 | 1 | 1 | 0 | 0 |
| batch19 | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 0 |
| batch20 | 4 | 3 | 6 | 1 | 1 | 0 | 0 | 0 |
| batch21 | 4 | 3 | 6 | 2 | 0 | 0 | 0 | 0 |
| batch22 | 3 | 6 | 2 | 3 | 1 | 0 | 0 | 0 |
| batch23 | 5 | 1 | 5 | 1 | 1 | 2 | 0 | 0 |
| batch24 | 4 | 5 | 2 | 3 | 0 | 0 | 1 | 0 |
| batch25 | 4 | 3 | 2 | 3 | 0 | 1 | 2 | 0 |
| batch26 | 3 | 3 | 4 | 1 | 2 | 2 | 0 | 0 |
| batch27 | 2 | 3 | 6 | 2 | 1 | 1 | 0 | 0 |
| batch28 | 3 | 3 | 4 | 4 | 0 | 0 | 0 | 1 |
| batch29 | 3 | 2 | 3 | 6 | 0 | 0 | 1 | 0 |
| batch30 | 3 | 1 | 5 | 4 | 1 | 1 | 0 | 0 |
| batch31 | 5 | 2 | 3 | 3 | 1 | 0 | 0 | 1 |
| batch32 | 4 | 4 | 4 | 0 | 3 | 0 | 0 | 0 |
| batch33 | 3 | 7 | 1 | 3 | 0 | 1 | 0 | 0 |
| batch34 | 3 | 4 | 5 | 2 | 0 | 1 | 0 | 0 |
| batch35 | 6 | 2 | 3 | 0 | 3 | 0 | 1 | 0 |
| batch36 | 2 | 5 | 3 | 3 | 1 | 1 | 0 | 0 |
| batch37 | 3 | 1 | 4 | 2 | 3 | 2 | 0 | 0 |
| batch38 | 2 | 4 | 3 | 3 | 1 | 2 | 0 | 0 |
| batch39 | 3 | 4 | 4 | 3 | 1 | 0 | 0 | 0 |
| batch40 | 3 | 2 | 1 | 4 | 2 | 3 | 0 | 0 |
| batch41 | 6 | 4 | 2 | 1 | 2 | 0 | 0 | 0 |
| batch42 | 5 | 3 | 2 | 2 | 1 | 2 | 0 | 0 |
| batch43 | 3 | 3 | 4 | 2 | 3 | 0 | 0 | 0 |
| batch44 | 2 | 2 | 5 | 2 | 2 | 1 | 1 | 0 |
| batch45 | 2 | 5 | 5 | 1 | 1 | 0 | 1 | 0 |
| batch46 | 2 | 5 | 2 | 3 | 3 | 0 | 0 | 0 |
| batch47 | 5 | 3 | 3 | 3 | 0 | 1 | 0 | 0 |
| batch48 | 4 | 3 | 5 | 1 | 2 | 0 | 0 | 0 |
| batch49 | 2 | 4 | 6 | 0 | 1 | 2 | 0 | 0 |
| batch50 | 2 | 2 | 4 | 1 | 5 | 1 | 0 | 0 |
| batch51 | 2 | 5 | 5 | 2 | 1 | 0 | 0 | 0 |
| batch52 | 1 | 3 | 4 | 3 | 4 | 0 | 0 | 0 |
| batch53 | 1 | 6 | 4 | 2 | 2 | 0 | 0 | 0 |
| batch54 | 3 | 4 | 4 | 2 | 2 | 0 | 0 | 0 |
| batch55 | 6 | 3 | 0 | 2 | 2 | 1 | 1 | 0 |
| batch56 | 1 | 6 | 1 | 5 | 2 | 0 | 0 | 0 |
| batch57 | 4 | 2 | 3 | 4 | 2 | 0 | 0 | 0 |
| batch58 | 1 | 3 | 6 | 2 | 3 | 0 | 0 | 0 |
| batch59 | 5 | 4 | 1 | 1 | 2 | 2 | 0 | 0 |
| batch60 | 4 | 3 | 1 | 3 | 2 | 1 | 1 | 0 |
| batch61 | 4 | 4 | 3 | 2 | 2 | 0 | 0 | 0 |
| batch62 | 5 | 5 | 4 | 1 | 0 | 0 | 0 | 0 |
| batch63 | 3 | 6 | 4 | 1 | 1 | 0 | 0 | 0 |
| batch64 | 2 | 3 | 6 | 3 | 0 | 1 | 0 | 0 |
| batch65 | 4 | 2 | 5 | 3 | 0 | 1 | 0 | 0 |
| batch66 | 2 | 5 | 4 | 4 | 0 | 0 | 0 | 0 |
| batch67 | 1 | 3 | 3 | 6 | 2 | 0 | 0 | 0 |
| batch68 | 2 | 3 | 6 | 3 | 0 | 1 | 0 | 0 |
| batch69 | 5 | 5 | 2 | 2 | 1 | 0 | 0 | 0 |
| batch70 | 4 | 3 | 4 | 2 | 2 | 0 | 0 | 0 |
| batch71 | 7 | 2 | 4 | 1 | 0 | 1 | 0 | 0 |
| batch72 | 1 | 3 | 3 | 3 | 4 | 1 | 0 | 0 |
| batch73 | 3 | 5 | 4 | 2 | 1 | 0 | 0 | 0 |
| batch74 | 6 | 2 | 3 | 1 | 1 | 1 | 1 | 0 |
| batch75 | 4 | 2 | 3 | 3 | 2 | 1 | 0 | 0 |
| batch76 | 3 | 2 | 6 | 2 | 1 | 1 | 0 | 0 |
| batch77 | 4 | 2 | 4 | 2 | 0 | 1 | 2 | 0 |
| batch78 | 3 | 2 | 5 | 2 | 1 | 1 | 1 | 0 |
| batch79 | 4 | 6 | 3 | 1 | 0 | 1 | 0 | 0 |
| batch80 | 4 | 5 | 1 | 2 | 2 | 1 | 0 | 0 |

## Factor 3  
prop\_table <- prop.table(factor3\_batch\_table, margin = 2)  
# Convert to a data frame for ggplot2  
prop\_df <- as.data.frame(as.table(prop\_table))  
colnames(prop\_df) <- c("Factor3", "Batch", "Proportion")  
  
# Create the stacked bar plot  
f3 <- ggplot(prop\_df, aes(x = Batch, y = Proportion, fill = Factor3)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Proportion of Age Across Batches", x = "Batch", y = "Proportion") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))  
f3

A colorful chart with numbers

Description automatically generated with medium confidence

# 4. Check distribution of factor4 across batches  
factor4\_batch\_table <- table(df$factor4, df$batchID)  
# print(factor4\_batch\_table)  
knitr::kable(t(factor4\_batch\_table), caption = "Distribution of Factor 4 Across Batches", format = "markdown")

Distribution of Factor 4 Across Batches

|  | Black | Other | White |
| --- | --- | --- | --- |
| batch1 | 1 | 0 | 14 |
| batch2 | 4 | 0 | 11 |
| batch3 | 1 | 0 | 14 |
| batch4 | 4 | 0 | 11 |
| batch5 | 2 | 0 | 13 |
| batch6 | 1 | 0 | 14 |
| batch7 | 2 | 1 | 12 |
| batch8 | 1 | 2 | 12 |
| batch9 | 3 | 0 | 12 |
| batch10 | 3 | 1 | 11 |
| batch11 | 3 | 0 | 12 |
| batch12 | 1 | 1 | 13 |
| batch13 | 4 | 0 | 11 |
| batch14 | 6 | 0 | 9 |
| batch15 | 3 | 1 | 11 |
| batch16 | 3 | 0 | 12 |
| batch17 | 2 | 1 | 12 |
| batch18 | 1 | 1 | 13 |
| batch19 | 3 | 0 | 12 |
| batch20 | 2 | 0 | 13 |
| batch21 | 4 | 0 | 11 |
| batch22 | 1 | 0 | 14 |
| batch23 | 2 | 2 | 11 |
| batch24 | 3 | 1 | 11 |
| batch25 | 1 | 0 | 14 |
| batch26 | 3 | 0 | 12 |
| batch27 | 3 | 0 | 12 |
| batch28 | 3 | 0 | 12 |
| batch29 | 2 | 0 | 13 |
| batch30 | 3 | 1 | 11 |
| batch31 | 2 | 0 | 13 |
| batch32 | 1 | 1 | 13 |
| batch33 | 1 | 1 | 13 |
| batch34 | 3 | 1 | 11 |
| batch35 | 1 | 0 | 14 |
| batch36 | 2 | 0 | 13 |
| batch37 | 1 | 0 | 14 |
| batch38 | 1 | 1 | 13 |
| batch39 | 3 | 0 | 12 |
| batch40 | 2 | 0 | 13 |
| batch41 | 2 | 0 | 13 |
| batch42 | 2 | 0 | 13 |
| batch43 | 3 | 0 | 12 |
| batch44 | 1 | 1 | 13 |
| batch45 | 2 | 0 | 13 |
| batch46 | 1 | 1 | 13 |
| batch47 | 3 | 0 | 12 |
| batch48 | 2 | 0 | 13 |
| batch49 | 1 | 0 | 14 |
| batch50 | 2 | 0 | 13 |
| batch51 | 3 | 1 | 11 |
| batch52 | 2 | 0 | 13 |
| batch53 | 1 | 0 | 14 |
| batch54 | 1 | 0 | 14 |
| batch55 | 3 | 0 | 12 |
| batch56 | 2 | 0 | 13 |
| batch57 | 5 | 0 | 10 |
| batch58 | 2 | 0 | 13 |
| batch59 | 1 | 1 | 13 |
| batch60 | 1 | 0 | 14 |
| batch61 | 2 | 0 | 13 |
| batch62 | 1 | 1 | 13 |
| batch63 | 4 | 0 | 11 |
| batch64 | 2 | 1 | 12 |
| batch65 | 3 | 1 | 11 |
| batch66 | 2 | 0 | 13 |
| batch67 | 1 | 1 | 13 |
| batch68 | 3 | 1 | 11 |
| batch69 | 4 | 0 | 11 |
| batch70 | 4 | 0 | 11 |
| batch71 | 1 | 2 | 12 |
| batch72 | 3 | 0 | 12 |
| batch73 | 4 | 0 | 11 |
| batch74 | 3 | 1 | 11 |
| batch75 | 1 | 1 | 13 |
| batch76 | 0 | 0 | 15 |
| batch77 | 2 | 0 | 13 |
| batch78 | 3 | 0 | 12 |
| batch79 | 2 | 0 | 13 |
| batch80 | 0 | 0 | 15 |

## Factor 4  
prop\_table <- prop.table(factor4\_batch\_table, margin = 2)  
# Convert to a data frame for ggplot2  
prop\_df <- as.data.frame(as.table(prop\_table))  
colnames(prop\_df) <- c("Factor4", "Batch", "Proportion")  
  
# Create the stacked bar plot  
f4 <- ggplot(prop\_df, aes(x = Batch, y = Proportion, fill = Factor4)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Proportion of Race Across Batches", x = "Batch", y = "Proportion") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))  
f4

A graph of a graph

Description automatically generated with medium confidence