

PCA__new

Jieqi Tu, jt3098

11/30/2019

```
# Import three datasets
CL = readxl::read_excel("./ABC_Cord Blood_Metabolomics_CL data_14Nov2019.xlsx") %>% janitor::clean_names()
BA = readxl::read_excel("./ABC_Cord Blood_Metabolomics_BA data_14Nov2019.xlsx") %>% janitor::clean_names()
PM = readxl::read_excel("./ABC_Cord Blood_Metabolomics_PM data_14Nov2019.xlsx") %>% janitor::clean_names()

# Extract only the data part of each dataset
CL_data = CL[11:491]
BA_data = BA[11:266]
PM_data = PM[11:193]

# Calculate the minimum value of each column
CL_data[CL_data == 0] = NA
BA_data[BA_data == 0] = NA
PM_data[PM_data == 0] = NA
CL_min = sapply(CL_data[1:481], function(x) min(x, na.rm = T))
BA_min = sapply(BA_data[1:256], function(x) min(x, na.rm = T))
PM_min = sapply(PM_data[1:183], function(x) min(x, na.rm = T))
CL_data = CL[11:491]
BA_data = BA[11:266]
PM_data = PM[11:193]

# Convert 0 to half of the minimum value
for(i in 1:481) {
  CL_data[i][CL_data[i]==0] = 0.5*CL_min[i]
}

for(i in 1:256) {
  BA_data[i][BA_data[i]==0] = 0.5*BA_min[i]
}

for(i in 1:183) {
  PM_data[i][PM_data[i]==0] = 0.5*PM_min[i]
}
```