

Data Exploration

Yiming Miao & Kexin Wang

Load dataset and libraries

```
library(ggplot2)
data <- readRDS("data/ukbiobank.rds")
```

Histogram of date of PD reporting

```
date_PD <- data$date_of_parkinsons_disease_report_f42032_0_0
df_date <- data.frame(date_PD = as.Date(date_PD[!is.na(date_PD)]), format = "%Y-%m-%d")
df_date$year_PD <- format(df_date$date_PD, "%Y")

ggplot(df_date, aes(x = year_PD)) +
  geom_bar(fill = "lightblue") +
  labs(title = "Histogram of PD Reporting Year",
       x = "Reporting Year", y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 7))
```

