

```

---
title: "hpc3"
author: "Jonathan Talbot"
date: "9/30/2021"
output:
  word_document: default
  pdf_document: default
  html_document: default
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```{r packages}
install.packages("tidyr")
library(tidyverse)
install.packages("readr")
library(readr)
install.packages("rio")
library(rio)
install.packages("sqldf")
library(sqldf)
install.packages("lubridate")
library(lubridate)
install.packages("ggplot2")
library(ggplot2)
install.packages("magrittr")
library(magrittr)
```

```{r load}
df <-data.frame()
df <- read.delim("household_power_consumption.txt", sep = ";") #load file from
local drive
df<-as_tibble(df)
print(head(df,100))
```

```{r lubridate}
df$Date <-lubridate::dmy(df$Date)
df$Time <-lubridate::hms(df$Time)
head(df)
```

```{r convert}
df[is.na(df)] <- 0 #replace NAs with 0s
any(is.na(df)) # check for any NAs remaining - should return False
df$Global_active_power<-as.numeric(df$Global_active_power) #"Numeric" #change
class of Global_active_power column from character to numeric
df$Global_reactive_power<-as.numeric(df$Global_reactive_power)
df$Voltage<-as.numeric(df$Voltage)
df$Global_intensity<-as.numeric(df$Global_intensity)
df$Sub_metering_1<-as.numeric(df$Sub_metering_1)

```

```

df$Sub_metering_2<-as.numeric(df$Sub_metering_2)
df$Sub_metering_3<-as.numeric(df$Sub_metering_3)
print(head(df,100))
```

```{r write}
#convert .txt to .csv
library("sqldf")
write.csv(df,"household_power_consumption.csv", row.names = TRUE)
print(summary(df))
```

```{r limit}

#limit to relevant dates
print("number of rows of df:")
print(nrow(df))
df_limited <- filter(df, Date == "2007-02-01" | Date == "2007-02-02")
print("number of rows of df_limited:")
print(nrow(df_limited))
print(head(df_limited))

```

```{r mutate}

#df_limited %>%
#  select(Date, Time, Global_active_power, Global_reactive_power, Voltage,
Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) #%>%
df_limited<- mutate(df_limited, Datetime = make_datetime(Date, Time))

print(head(df_limited))
```

```{r mutate 2}
df_limited<-mutate(df_limited, Weekday = wday(Date))
print(head(df_limited))
```

```{r}
print(tail(df_limited))
```

```{r plot1}

#create plot
png("Rplot1.png",width = 480, height = 480, units = "px", pointsize = 12)

myplot<-ggplot(data=df_limited, mapping=aes(x=Global_active_power ))+
geom_histogram(color="orange", fill="dark orange")+
  ylab("Frequency")+

```

```

    xlab("Global active power (kilowatts)")

#save as .png 480 x 480

print(myplot)
dev.off()

...

```{r plot2}
print(head(df_limited))
#Plot number 2
library(ggplot2)
#png("Rplot2.png",width = 480, height = 480, units = "px", pointsize = 12)

#myplot2 <-

qplot(df_limited$Datetime, df_limited$Global_active_power)

#myplot2<-ggplot(data=df_limited,
mapping=aes(x=Datetime,y=Global_active_power))+
#      geom_point() #+
#      #ylab("Global Active Power (kilowatts)")
#)

#print(myplot2)
#dev.off()

...

```{r plot3}
#Plot number 3 - submetering

#png("Rplot3.png", width = 480, height = 480, units = "px", pointsize = 12)
plot(df_limited$Datetime, df_limited$Sub_metering_1, col = "black")
plot(df_limited$Datetime, df_limited$Sub_metering_2, col = "red")
plot(df_limited$Datetime, df_limited$Sub_metering_3, col = "blue", ylab =
"Energy sub metering")

#myplot3 <- ggplot(data=df_limited, mapping=aes(x=Datetime)) +
#  geom_line(aes(y = Sub_metering_1), color = "black") +
#  geom_line(aes(y = Sub_metering_2), color="red")+
#  geom_line(aes(y = Sub_metering_3), color="blue")

#print(myplot3)
#dev.off()

...

```

```
```{r plot4}
#Plot number 4
#myplot4 <-png(filename = "Rplot4.png",
#      width = 480, height = 480, units = "px", pointsize = 12)

qplot(y=df_limited$Global_reactive_power, x=df_limited$Datetime)

#print(myplot3)
#dev.off()
```
```