Untitled

ME

2023-04-08

{r setup, include=FALSE} knitr::opts\_chunk$set(echo = TRUE)

library(ggplot2)

## Loading and preprocessing the data

path=getwd() unzip(“repdata\_data\_activity.zip”, exdir=path) activity<- read.csv(“activity.csv”, header=TRUE, sep=‘,’,colClasses= c(“numeric”, “character”, “integer”)) activitydate, “%Y%m%d”) day<- weekdays(activity$date) activity<- cbind(activity, day) summary(activity)

## What is mean total number of steps taken per day?

activityTotalSteps <- with(activity, aggregate(steps, by = list(date), sum, na.rm = TRUE))

names(activityTotalSteps) <- c(“Date”, “Steps”)

totalStepsdf <- data.frame(activityTotalSteps)

g <- ggplot(totalStepsdf, aes(x = Steps)) + geom\_histogram(breaks = seq(0, 25000, by = 2500), fill = “#83CAFF”, col = “black”) + ylim(0, 30) + xlab(“Total Steps Taken Per Day”) + ylab(“Frequency”) + ggtitle(“Total Number of Steps Taken on a Day”)

print(g)

##Mean number of steps taken per day mean(activityTotalSteps$Steps)

##The median of the total number of steps taken per day is: median(activityTotalSteps$Steps)

## What is the average daily activity pattern?

averageDailyActivity <- aggregate(activityinterval), FUN = mean, na.rm = TRUE)

names(averageDailyActivity) <- c(“Interval”, “Mean”)

averageActivitydf <- data.frame(averageDailyActivity)

gg <- ggplot(averageActivitydf, mapping = aes(Interval, Mean)) + geom\_line(col = “darkred”) + xlab(“Interval”) + ylab(“Average Number of Steps”) + ggtitle(“Average Number of Steps Per Interval”)

print(gg)

## Imputing missing values

sum(is.na(activityMean[match(activityInterval)] activityImputed <- transform(activity, steps = ifelse(is.na(activitysteps))

totalActivityImputed <- aggregate(steps ~ date, activityImputed, sum)

names(totalActivityImputed) <- c(“date”, “dailySteps”) sum(is.na(totalActivityImputed$dailySteps)) totalImputedStepsdf <- data.frame(totalActivityImputed)

gg2 <- ggplot(totalImputedStepsdf, aes(x = dailySteps)) + geom\_histogram(breaks = seq(0, 25000, by = 2500), fill = “#83CAFF”, col = “black”) + ylim(0, 30) + xlab(“Total Steps Taken Per Day”) + ylab(“Frequency”) + ggtitle(“Total Number of Steps Taken on a Day”)

print(gg2) ##The mean of the total number of steps taken per day is:

mean(totalActivityImputed$dailySteps)

##The median of the total number of steps taken per day is:

median(totalActivityImputed$dailySteps)

## Are there differences in activity patterns between weekdays and weekends?

# Updating format of the dates

activitydate, format=“%Y-%m-%d”))

# Creating a function that distinguises weekdays from weekends

activitydate, function(x) { if(weekdays(x) == “Saturday” | weekdays(x) == “Sunday”) {y <- “Weekend”} else {y <- “Weekday”} y }) activityByDay <- aggregate(steps ~ interval + dayType, activity, mean, na.rm = TRUE)

# Plotting using ggplot2

dayPlot <- ggplot(activityByDay, aes(x = interval , y = steps, color = dayType)) + geom\_line() + ggtitle(“Average Daily Steps by Day Type”) + xlab(“Interval”) + ylab(“Average Number of Steps”) + facet\_wrap(~dayType, ncol = 1, nrow=2) + scale\_color\_discrete(name = “Day Type”)

print(dayPlot)