Reproducible research

Nicholas

2024-04-01

```
Activity <- read.csv("activity.csv")
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
Activity <-  as_tibble(Activity)
```

What is mean total number of steps taken per day?

```
Total_step_day<-Activity %>%group_by(date) %>%  summarise(steps=sum(steps,na.rm = TRUE))
png("plot11.png",width = 480,height=480)
hist(Total_step_day$steps,breaks = 100)
dev.off()
```

```
## pdf
##   2
```

```
mean(Total_step_day$steps)
```

```
## [1] 9354.23
```

```
median(Total_step_day$steps)
```

```
## [1] 10395
```

What is the average daily activity pattern?

```
Complete_Activity_int<-Activity %>%group_by(interval) %>% summarize(steps=mean(steps,na.rm=TRUE))
png("plot12.png",width = 480,height=480)
with(Complete_Activity_int,plot(interval,steps,type="l"))
dev.off()
```

```
## pdf
##   2
```

```
Complete_Activity_int %>%group_by(interval) %>%arrange(-steps) %>% filter(steps==max(Complete_Activity_
```

```
## # A tibble: 1 x 2
## # Groups:   interval [1]
##   interval steps
##      <int> <dbl>
## 1      835  206.
```

THe 835 interval has the highest average steps

Imputing missing values

```
Activity %>%count(steps %in% NA)
```

```
## # A tibble: 2 x 2
##   `steps %in% NA`     n
##   <lgl>           <int>
## 1 FALSE           15264
## 2 TRUE             2304
```

2304 missing values

filling in all of the missing values with the mean

```
mean_activity<- mean(Activity$steps,na.rm=TRUE)
Activity2<-Activity %>% mutate(steps=ifelse(steps %in% NA,mean_activity,steps))

Total_step_day2<-Activity2 %>% group_by(date) %>% summarise(steps=sum(steps))
png("plot13.png",width = 480,height=480)
hist(Total_step_day2$steps,breaks = 100)
dev.off()
```

```
## pdf
##   2
```

```
mean(Total_step_day2$steps)
```

```
## [1] 10766.19
```

```
median(Total_step_day2$steps)
```

```
## [1] 10766.19
```

Are there differences in activity patterns between weekdays and weekends?

```r
Activity2$date<- as.Date(Activity2$date)

Activity_week<- Activity2 %>% mutate(date=if_else((wday(Activity2$date) %in% 2:6),"Weekday","Weekend"))

Activity_week$date<- as.factor(Activity_week$date)

Activity_weekdays<-Activity_week %>%filter(date=="Weekday") %>% group_by(interval) %>%  summarise(steps=

Activity_weekends<-Activity_week %>% filter(date=="Weekend") %>%  group_by(interval) %>% summarise(steps

png("plot14.png",width = 480,height=480)
plot(Activity_weekdays$interval,Activity_weekdays$steps,type = "l",xlab="interval",ylab="Average steps")
lines(Activity_weekends$interval,Activity_weekends$steps,col="red")
legend("topleft",legend=c("Weekday","Weekend"),
col=c("black","red"),lty=1)
dev.off()
```

```
## pdf
##   2
```