Project 3

Huiyue Li, Yifeng Tang, Yiyuan Yao<-Huiyue Li

```r
# load the packages we need to use
library(tidyverse)
library(knitr)
library(here)
library(lubridate)
# using the here package
here::here()

## [1] "/Users/heileylee/R scriptfile/Li_Project3"
```

Plotting

```r
# input the dataset
plot1<-read_csv("plot_data.csv")
# extract the month of each observation and select the relevant columns
plot2<-plot1%>%
  mutate(month=month(ymd(plot1$week)))%>%
  select(3:13)
# reorganize the dataframe by gathering the site, then calculate the me
an by month and site since reorganize the dataframe by gathering the si
te
# calculate the mean by month and site to represent the number of subje
ct enrolled by month and site, since the total number of times collecti
ng data varies among different months
pp=plot2%>%
  gather(Site,Numb,-month)%>%
  group_by(month,Site)%>%
  summarise(mean=mean(Numb))%>%arrange(Site)
# create a plot illustrating the number of subjects enrolled by month a
nd site
plotting<-ggplot(pp, aes(x = factor(month), y = mean,color=factor(Site,
level=c("Site1","Site2","Site3","Site4","Site5","Site6","Site7","Site8",
"Site9","Site10")))) +
  geom_point(size=3.2,alpha=.75) +
  geom_text(aes(label=round(mean,2)),nudge_y=.5,color="black",size=2.6,
check_overlap = T)+
  scale_colour_discrete(name="Site",labels=c("Site 1","Site 2","Site 3",
"Site 4","Site 5","Site 6","Site 7","Site 8","Site 9","Site 10"))+
  xlab("Month") + ylab("Numberof enrollment") +
  ggtitle("# The number of subjects enrolled by month and site")+
  theme_bw()+theme(legend.position = "bottom")+
  facet_wrap(~factor(Site,level=c("Site1","Site2","Site3","Site4","Site
5","Site6","Site7","Site8","Site9","Site10")),nrow = 5,ncol = 2,scales
= "free_x")
plotting
```

The number of subjects enrolled by month and site
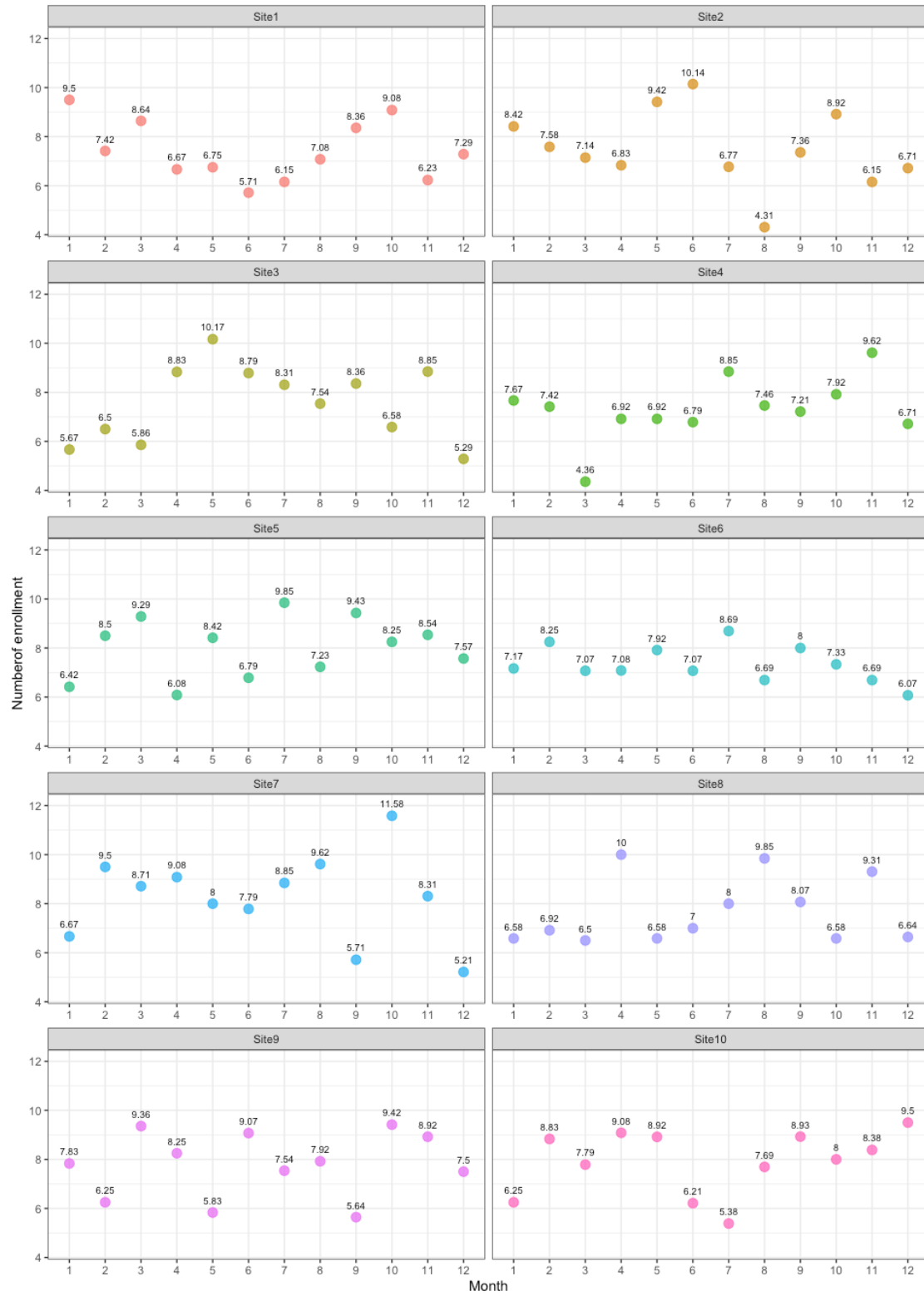
Program control

```r
# input the dataset
hiv<-read.csv("HIV_data.csv")
# create an empty vector for loop results
hiv1<-NULL
# create a FOR loop to detect the whole dataframe except the column 1
# determine if there is a "CD4 count<400" using an IF statement
for(i in 1:nrow(hiv)){
  for(j in 2:6){
  if(hiv[i,j]<400){
  hiv1=rbind(hiv1,c(i,paste("V",j-1,sep=""),hiv[i,j]))
  }
  }
}
# transform hiv1 into a dataframe
hiv1<-data.frame(hiv1)

# divide the hiv1 dataframe into two parallel groups to avoid page cut
# the first 26 rows in left part
hiv_1=hiv1[1:26,]
# the 26th~51th rows in right part
hiv_2=hiv1[27:51,]
# set a vector with NA
co=c('', '', '')
# complement the right part(keeping left and right rows the same)
hiv_3=rbind(hiv_2,co)
# bind the columns all together
hiv_final<- cbind(hiv_1,bre=rep('|',26),hiv_3)
# see the result, i.e. the kable, on the next page
```

```r
# hide the NA in the kable
options(knitr.kable.NA = '')
# display the result by kable
kable(hiv_final,caption = "The patient ID and study visit for patients
with CD4 smaller than 400",align = "c",col.names = c("**Patient ID**","
**Study Visit**","**CD4 Count**","**|**","**Patient ID**","**Study Visi
t**","**CD4 Count**"))
```

The patient ID and study visit for patients with CD4 smaller than 400

Patient ID	Study Visit	CD4 Count			Patient ID	Study Visit	CD4 Count
3	V5	346	|	248	V2	144
10	V1	387	|	253	V3	173
12	V4	374	|	256	V5	328
15	V2	360	|	265	V1	371
28	V1	339	|	272	V1	390
31	V3	283	|	290	V2	292
44	V4	288	|	304	V4	399
47	V5	284	|	321	V5	327
50	V3	363	|	329	V5	395
73	V4	366	|	330	V4	392
104	V2	342	|	337	V5	345
107	V2	392	|	340	V1	372
122	V5	385	|	346	V2	138
133	V1	361	|	369	V3	354
135	V3	307	|	383	V3	373
141	V5	388	|	384	V5	387
151	V5	388	|	397	V2	393
152	V1	338	|	414	V5	366
166	V2	379	|	425	V4	382
183	V2	297	|	426	V3	396
185	V5	390	|	446	V1	326
228	V2	359	|	450	V3	333
234	V3	289	|	457	V2	329
242	V2	266	|	465	V1	288
243	V4	80	|	478	V2	390
247	V5	292	|			