HW4_fish_data

Ruoyuan Li

February 16, 2018

BMB: it's good to name all your R code chunks.

First load package and data.

```r
library(readr)
library(reshape2) ## BMB: why not tidyr?
library(tidyverse)
library(directlabels)

fish_data <- read_csv("fish.csv",col_names = TRUE)
colnames(fish_data)[3] <- "SampleID"
## BMB: could also use names(fish_data) <- make.names(fish_data)
##  or dplyr::rename()
## BMB: col_names=TRUE is the default value ...
fish_name <- read_csv("fish_names.csv",col_names = TRUE)
```

Switch wide format to long format, and merge the two data sets together. Now the full data has all the information for plotting.

```r
#wide format to long format
fish_long <- melt(fish_data, id.vars = c("SampleID", "Site","MetCode","SamplerType"),
    variable.name = "treatment_variable",
    value.name = "treatment_value")

#merge name data with  long format data.
fish_name$treatment_variable<- fish_name$abbr
full<-merge(fish_long, fish_name)
(full%>%select(-abbr))->full
```

First plot showing under the same drugcat, how drug changes with Site. Like under drugcat is food, SUC always have higher value for each Site than CFN, same happen to antidepressant, and antibiotic. However, like with drugcate is beta-blocker, ATN has higher value than PPN for both Site: Downstream and Outfall, but PPN higher than ATN for site:WWTP.

```r
p1<-(ggplot(full,aes(x=Site,y=treatment_value,colour=treatment_variable))
    +geom_point(alpha=0.3)
    +facet_wrap(~drugcat,labeller = label_both,scale = "free_y"))

print(p1+geom_line(aes(group=treatment_variable))
   +geom_dl(aes(label=treatment_variable),,method="last.points")
   +theme(axis.text.x = element_text(angle = 60, hjust=1))
   +theme(legend.position="none")
)
```
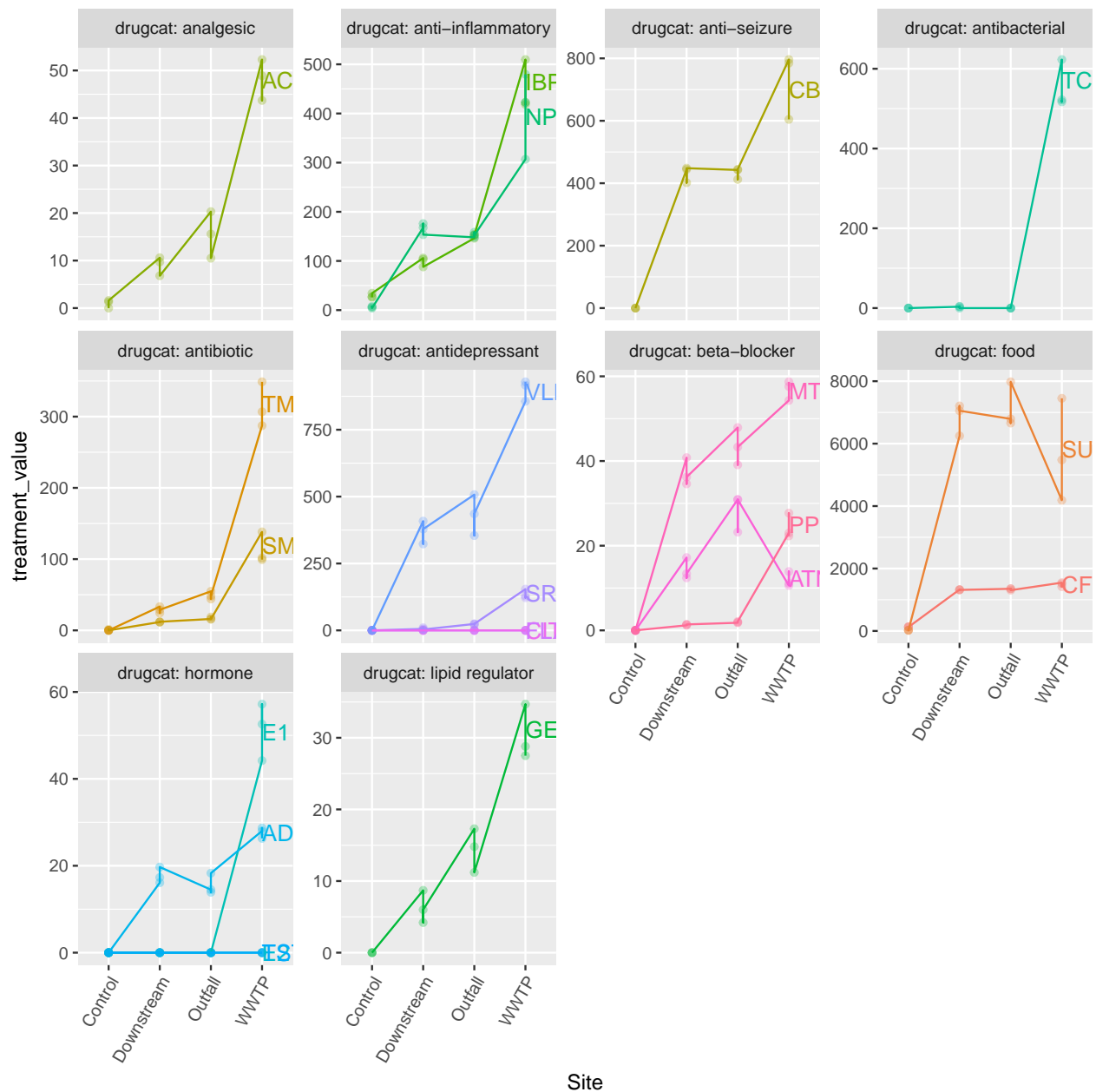
BMB: I adjusted the justificatino with `hjust` to align the x-axis labels more neatly, and adjusted the width and height of the figure (`fig.width` and `fig.height` chunk options)

Second plot is drug's value for each Site. SUC has much higher value for each Site except Control. Compared to onther Site, control almost has fewer changes for each drug.

```
p2<-ggplot(full,aes(x=treatment_variable,y=treatment_value,colour=Site))
print(p2+geom_line(aes(group=Site))+facet_wrap(~Site,labeller = label_both)
      +geom_point()
      +theme(axis.text.x = element_text(angle = 60))
      +theme(legend.position="none"))
```
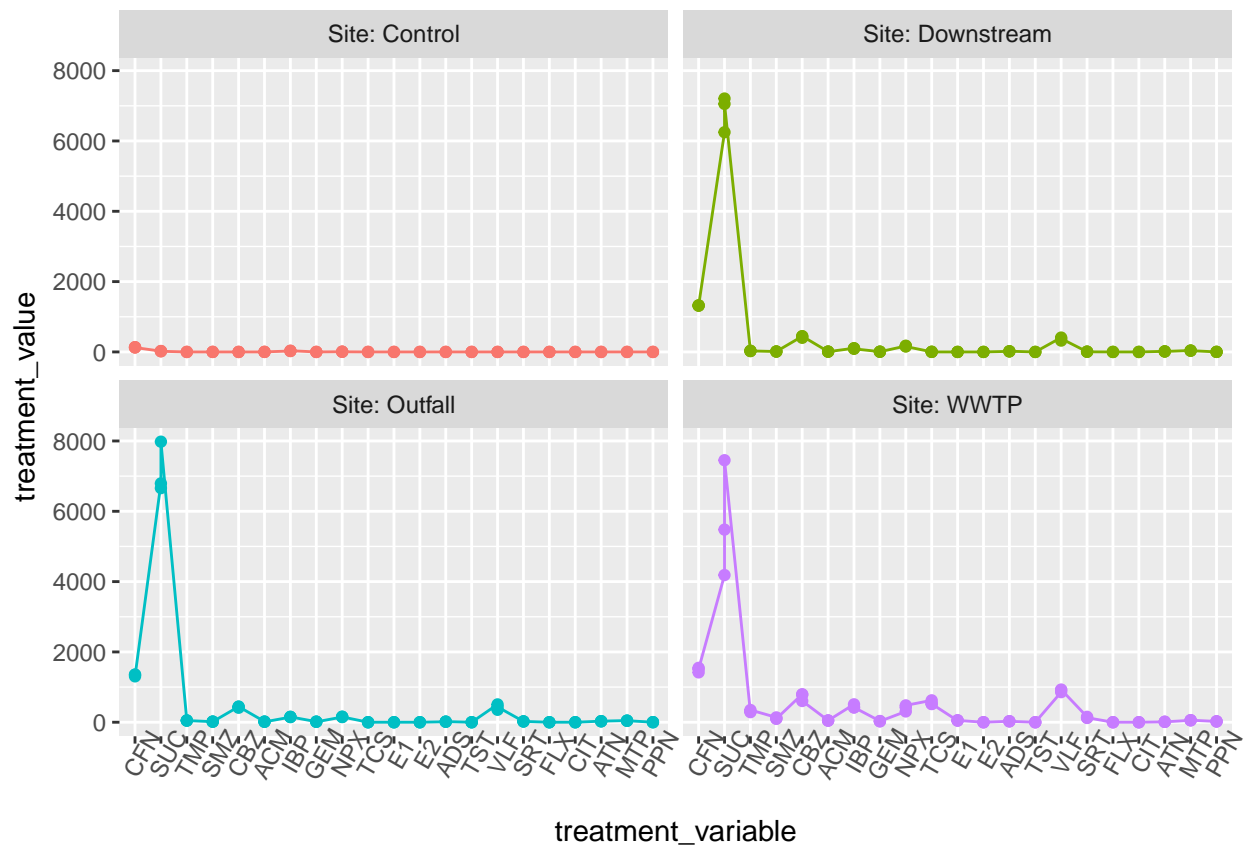
```
print(p2+geom_line(aes(group=Site))+facet_wrap(~Site,labeller = label_both,scales = "free")
      +geom_point()
      +theme(axis.text.x = element_text(angle = 60))
      +theme(legend.position="none"))
```