HW 5 Busayo

Olubusayo Bolonduro

8/31/2020

Tidying dad_mom

```
dad_mom <-read.table("dad_mom.txt", header = TRUE)
tidy <- dad_mom %>%
  unite(dad, name_dad, income_dad) %>%
  unite(mom, name_mom, income_mom)%>%
  gather(c(dad,mom), key="parent", value ="income")%>%
  separate(income, c("name","income"))
tidy
```

```
##   fam_id parent name income
## 1      1    dad Bill  30000
## 2      2    dad  Art  22000
## 3      3    dad Paul  25000
## 4      1    mom Bess  15000
## 5      2    mom  Amy  22000
## 6      3    mom  Pat  50000
```

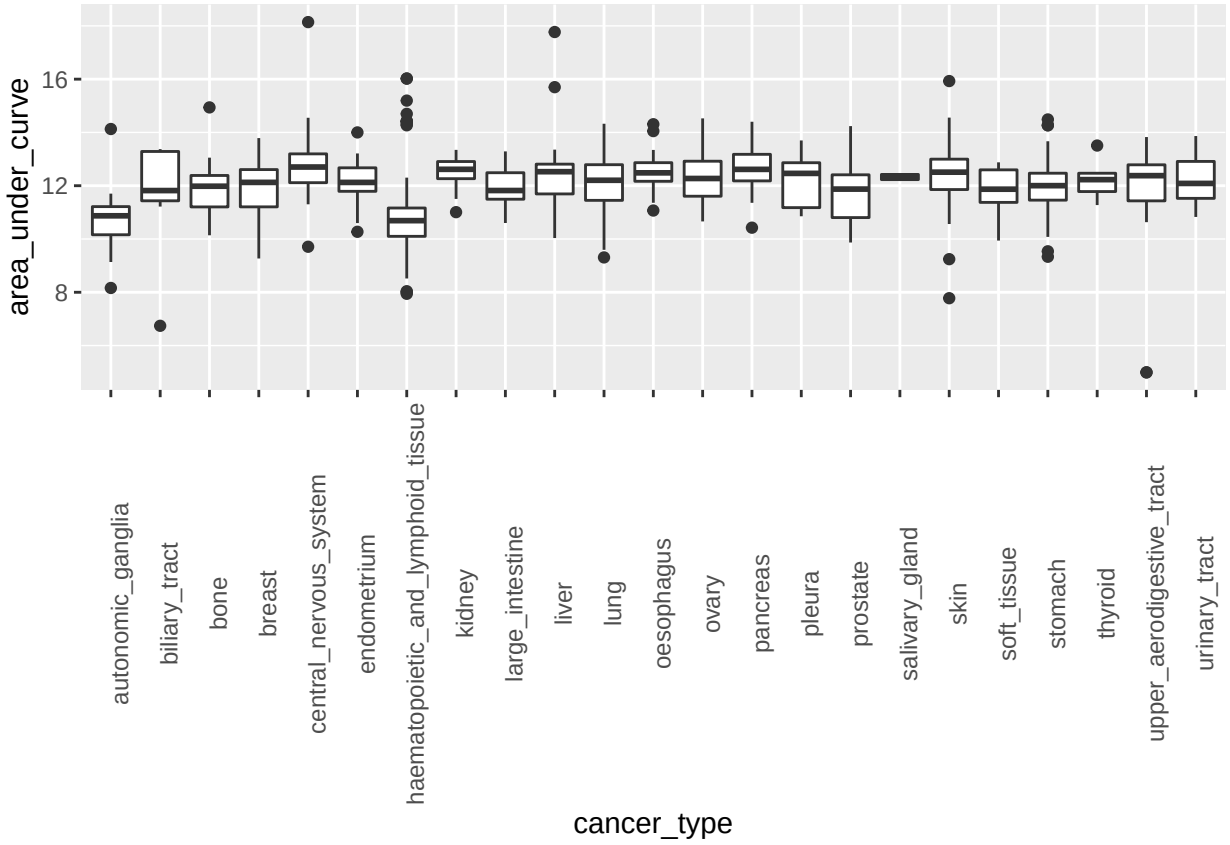
Joining CTRP data

```
auc1 <- read.csv("CTRP_files/AUC_1.csv", header = TRUE)
auc2 <- read.csv("CTRP_files/AUC_2.csv", header = TRUE)
ccl <- read.csv("CTRP_files/cancer_cell_line_info.csv", header = TRUE)
compound <- read.csv("CTRP_files/compound_info.csv", header = TRUE)
exp <- read.csv("CTRP_files/Experiment_info.csv", header = TRUE)

auc_merged <- auc1 %>%
  rbind(auc2)
All_data <- auc_merged%>%
  left_join(exp, by = c("experiment_id" = "expt_id"))%>%
  left_join(ccl, by = "master_ccl_id")%>%
  left_join(compound, by = "master_cpd_id")
#All_data
```

Q1 Which cancer type has the lowest AUC values to the compound "vorinostat"?

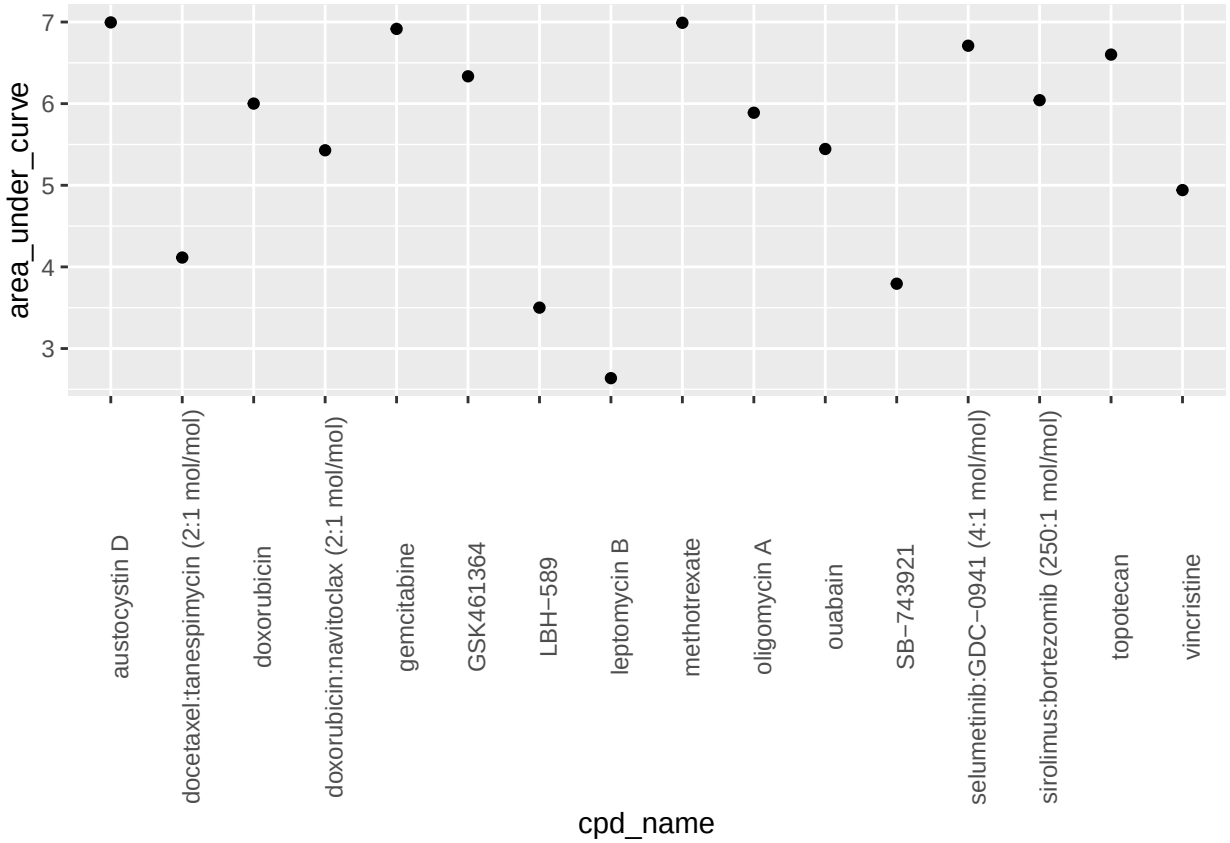
```
comp_vorinostat <- All_data %>%
  select(area_under_curve,cancer_type,cpd_name) %>%
  filter(cpd_name == "vorinostat")%>%
  group_by(cancer_type)%>%
  drop_na()
ggplot(comp_vorinostat, aes(cancer_type, area_under_curve))+
  geom_boxplot()+
  theme(axis.text.x = element_text(angle =90))
```




```
##The lowest vorinostat value appeared under the upper aerodigestive tract cancer type.
```

Q2 Which compound is the prostate cancer cell line 22RV1 most sensitive to? (For 22RV1, which compound has the lowest AUC value?)

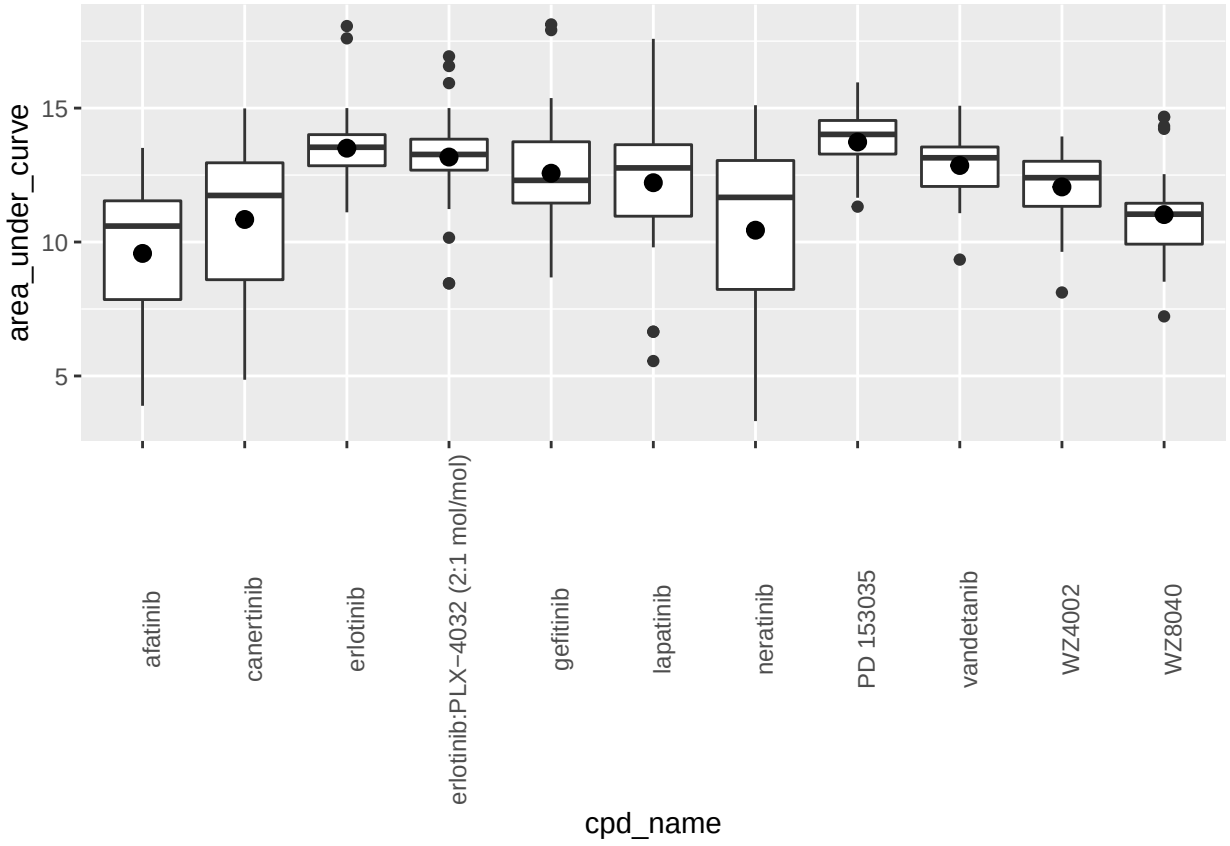
```
pcc <- All_data %>%
  select(area_under_curve,ccl_name,cpd_name)%>%
  ##I decided to filter for only lower AUC values
  filter(ccl_name=="22RV1", area_under_curve < 7)%>%
  group_by(cpd_name)%>%
  drop_na()
ggplot(pcc, aes(cpd_name, area_under_curve))+
  geom_point()+
  theme(axis.text.x = element_text(angle = 90))
```




```
## The compound with the lowest AUC value was leptomycin B.
```

Q3 For the 10 compounds that target EGFR, which of them has (on average) the lowest AUC values in the breast cancer cell lines?

```
egfr <- All_data %>%
  select(area_under_curve, cpd_name, gene_symbol_of_protein_target, cancer_type)%>%
  filter(cancer_type == "breast", grepl("EGFR", gene_symbol_of_protein_target))%>%
  group_by(cpd_name)%>%
  drop_na()
ggplot(egfr, aes(cpd_name, area_under_curve))+
  geom_boxplot()+
  #I found a quick way to overlay the mean on a boxplot
  stat_summary(fun = "mean")+
  theme(axis.text.x = element_text(angle = 90))
```




```
##The compound with the lowest average AUC in the breast cancer cell lines was afatinib.
```