

Exploratory Data Analysis

```
# EDA for (pre-post)/max values
```

```
# libraries
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3      v purrr  0.3.4
```

```
## v tibble  3.0.6      v dplyr  1.0.3
```

```
## v tidyr   1.1.2      v stringr 1.4.0
```

```
## v readr   1.4.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
library(gridExtra)
```

```
##
```

```
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      combine
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
library(knitr)
```

```
library(summarytools)
```

```
## Registered S3 method overwritten by 'pryr':
```

```
##   method      from
```

```
##   print.bytes Rcpp
```

```
## For best results, restart R session and update pander using devtools:: or remotes::install_github('r')
```

```
##
```

```
## Attaching package: 'summarytools'
```

```
## The following object is masked from 'package:tibble':
```

```
##
```

```
##      view
```

```

# load data
ss_all <-
  readRDS("Data/ss_all.RDS") %>%
  mutate(viewcat = as.factor(viewcat),
         sex = as.factor(sex),
         setting = as.factor(setting),
         encour = as.factor(encour),
         site = as.factor(site),
         regular = as.factor(regular))

# ggplot options
theme_set(theme_bw())

# cont. summary table
summary_num <-
  ss_all %>%
  select(-percbody, -perclet, -percnumb, -id) %>%
  descr(stats = c("mean", "sd", "min", "med", "max")) %>%
  tb()
kable(summary_num)

```

variable	mean	sd	min	med	max
age	51.52500	6.281357	34.00000	52.00000	69.00000
peabody	46.46667	16.038621	8.00000	42.00000	99.00000
percbody_max	12.05729	15.851114	-34.37500	12.50000	71.87500
perclet_max	18.62787	19.255401	-37.93103	15.51724	70.68966
percnumb_max	16.95216	17.954089	-64.81481	16.66667	61.11111
postbody	25.25833	5.500849	11.00000	27.00000	39.00000
postlet	26.74167	13.375176	0.00000	23.00000	63.00000
postnumb	30.00833	12.822572	0.00000	29.00000	54.00000
prebody	21.40000	6.390893	6.00000	22.00000	32.00000
prelet	15.93750	8.536425	1.00000	14.00000	55.00000
prenumb	20.85417	10.684893	1.00000	19.00000	52.00000

```

# cat. summary tables
freq(ss_all$site, cumul = F, report.nas = F, totals = F) %>% tb() %>% kable()

```

site	freq	pct
1	60	25.00000
2	55	22.91667
3	64	26.66667
4	43	17.91667
5	18	7.50000

```

freq(ss_all$sex, cumul = F, report.nas = F, totals = F) %>% tb() %>% kable()

```

sex	freq	pct
1	115	47.91667
2	125	52.08333

```
freq(ss_all$viewcat, cumul = F, report.nas = F, totals = F) %>% tb() %>% kable()
```

viewcat	freq	pct
1	54	22.50000
2	60	25.00000
3	64	26.66667
4	62	25.83333

```
freq(ss_all$setting, cumul = F, report.nas = F, totals = F) %>% tb() %>% kable()
```

setting	freq	pct
1	143	59.58333
2	97	40.41667

```
freq(ss_all$encour, cumul = F, report.nas = F, totals = F) %>% tb() %>% kable()
```

encour	freq	pct
0	88	36.66667
1	152	63.33333

```
# Not show how to make this table
# ss_all %>%
#   select(-c(regular, percnumb, perclet, percbody)) %>%
#   summary()
```

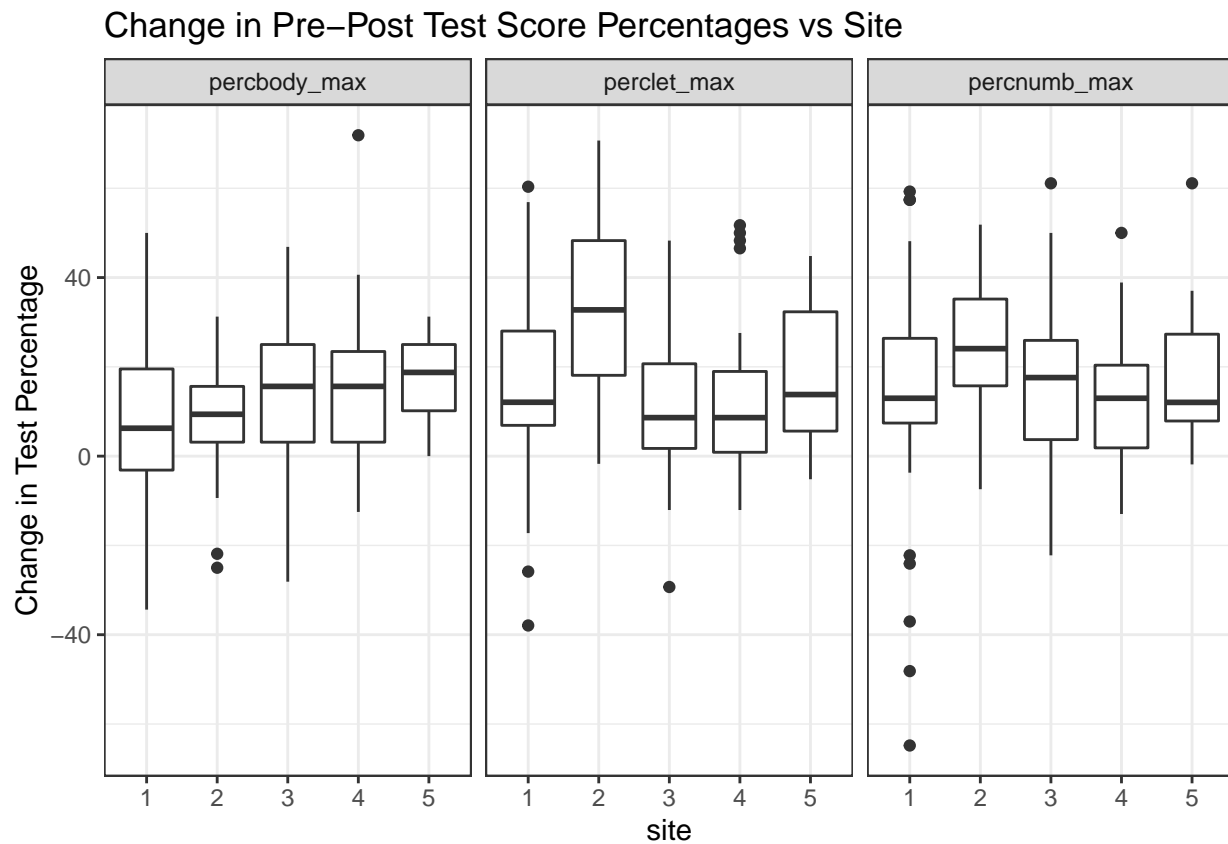
```
# viewcat boxplots
v1 <- ggplot(ss_all, aes(viewcat, percnumb_max)) +
  geom_boxplot() +
  ylim(-75, 75) +
  xlab("Viewing Category") +
  ylab("Change in Number Score Percentage")
v2 <- ggplot(ss_all, aes(viewcat, perclet_max)) +
  geom_boxplot() +
  ylim(-75, 75) +
  xlab("Viewing Category") +
  ylab("Change in Letter Score Percentage")
v3 <- ggplot(ss_all, aes(viewcat, percbody_max)) +
  geom_boxplot() +
  ylim(-75, 75) +
  xlab("Viewing Category") +
  ylab("Change in Body Score Percentage")
```

```
png("viewcat_boxplot.png", width = 1280, height = 720)
gridExtra::grid.arrange(v1, v2, v3, nrow = 1)
dev.off()
v1
v2
v3
```

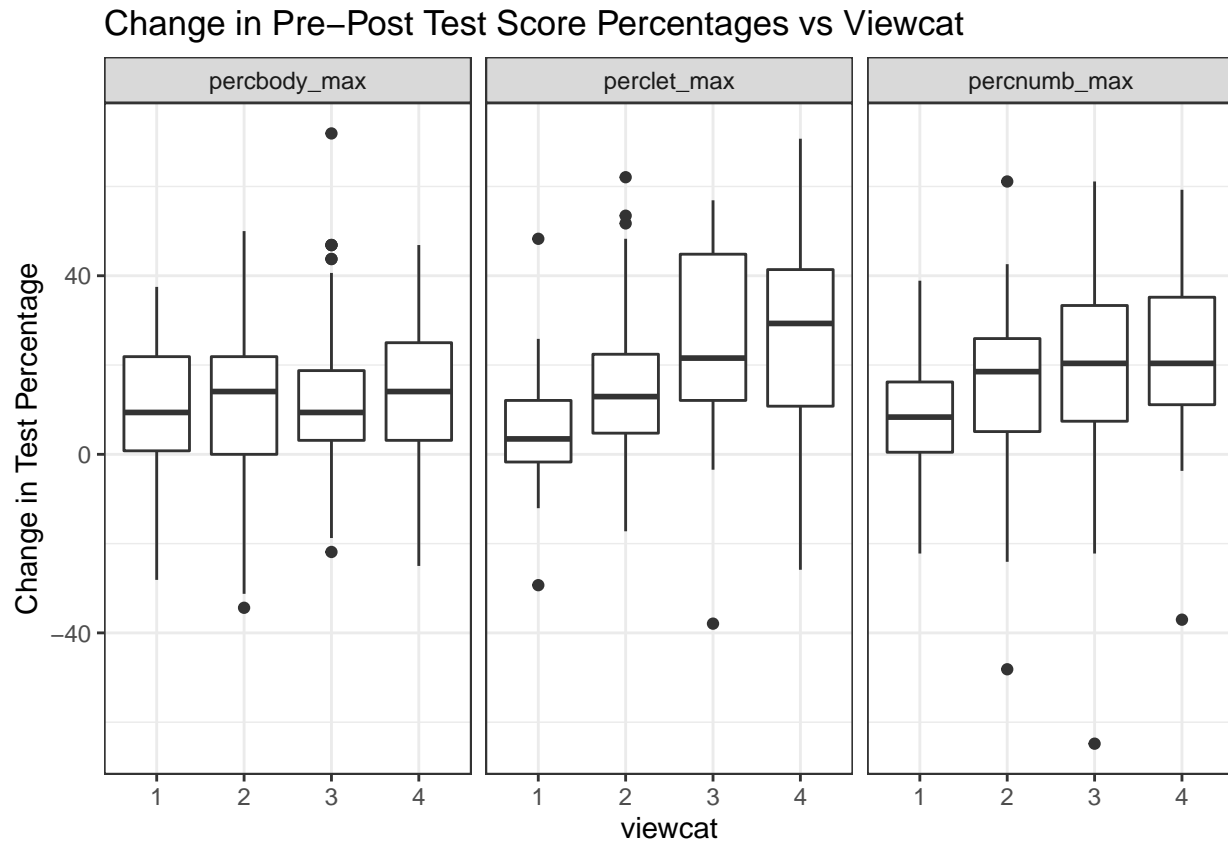
Boxplots for every categorical variable

```
# Convert the data from wide to long so we can put it one plot
ss_cat_long <-
  ss_all %>%
  select(id, percnumb_max, perclet_max, percbody_max,
         site, viewcat, sex, setting, encour) %>%
  pivot_longer(cols = c(percnumb_max, perclet_max, percbody_max))
```

```
ggplot(ss_cat_long, aes(x = site, y = value)) +
  geom_boxplot() +
  facet_wrap(~name) +
  labs(y = "Change in Test Percentage",
       title = "Change in Pre-Post Test Score Percentages vs Site")
```

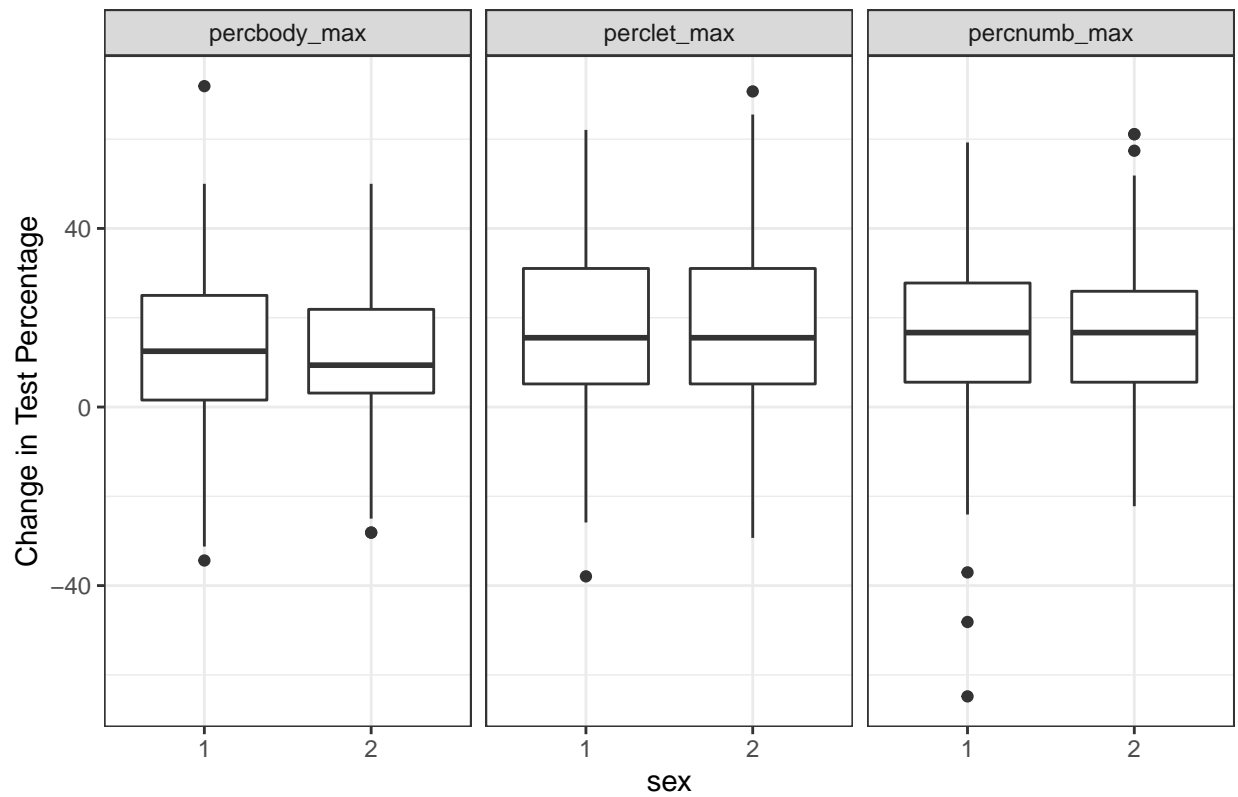


```
ggplot(ss_cat_long, aes(x = viewcat, y = value)) +
  geom_boxplot() +
  facet_wrap(~name) +
  labs(y = "Change in Test Percentage",
       title = "Change in Pre-Post Test Score Percentages vs Viewcat")
```



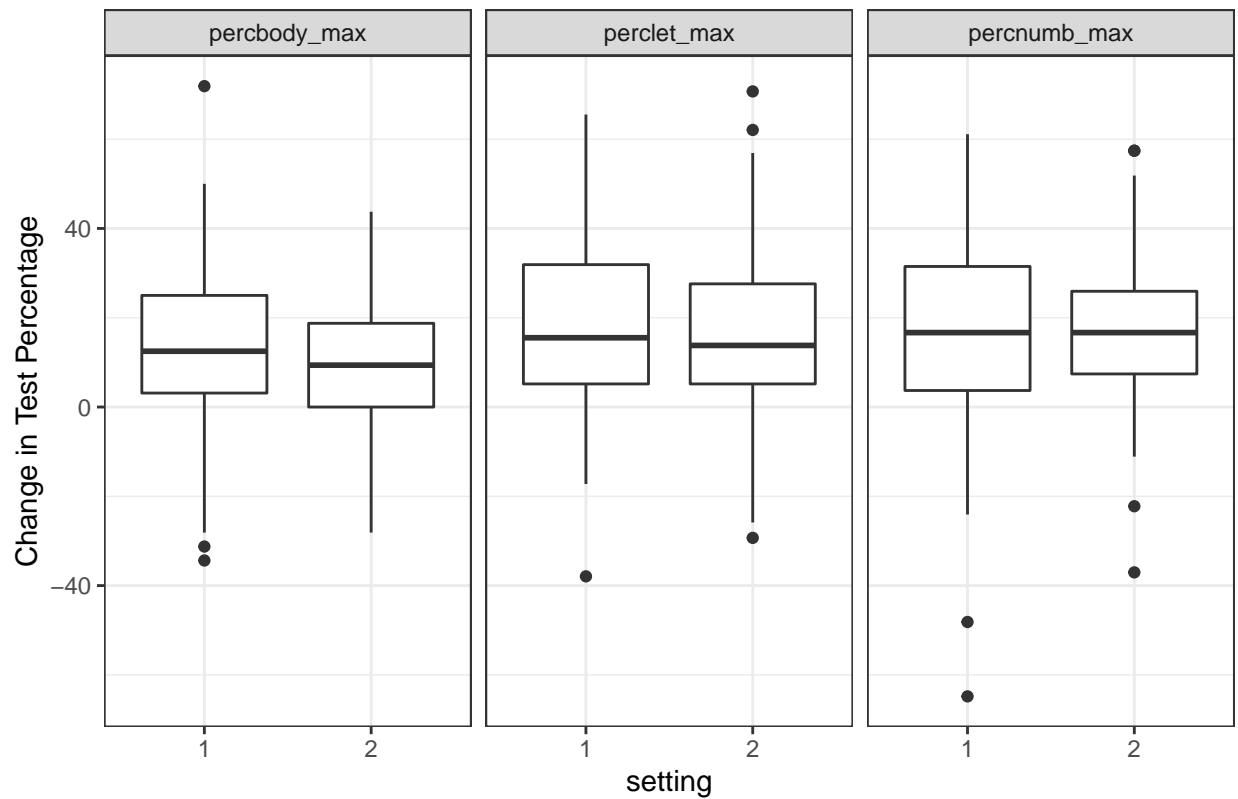
```
ggplot(ss_cat_long, aes(x = sex, y = value)) +
  geom_boxplot() +
  facet_wrap(~name) +
  labs(y = "Change in Test Percentage",
       title = "Child Change in Pre-Post Test Score Percentages vs Sex")
```

Child Change in Pre-Post Test Score Percentages vs Sex



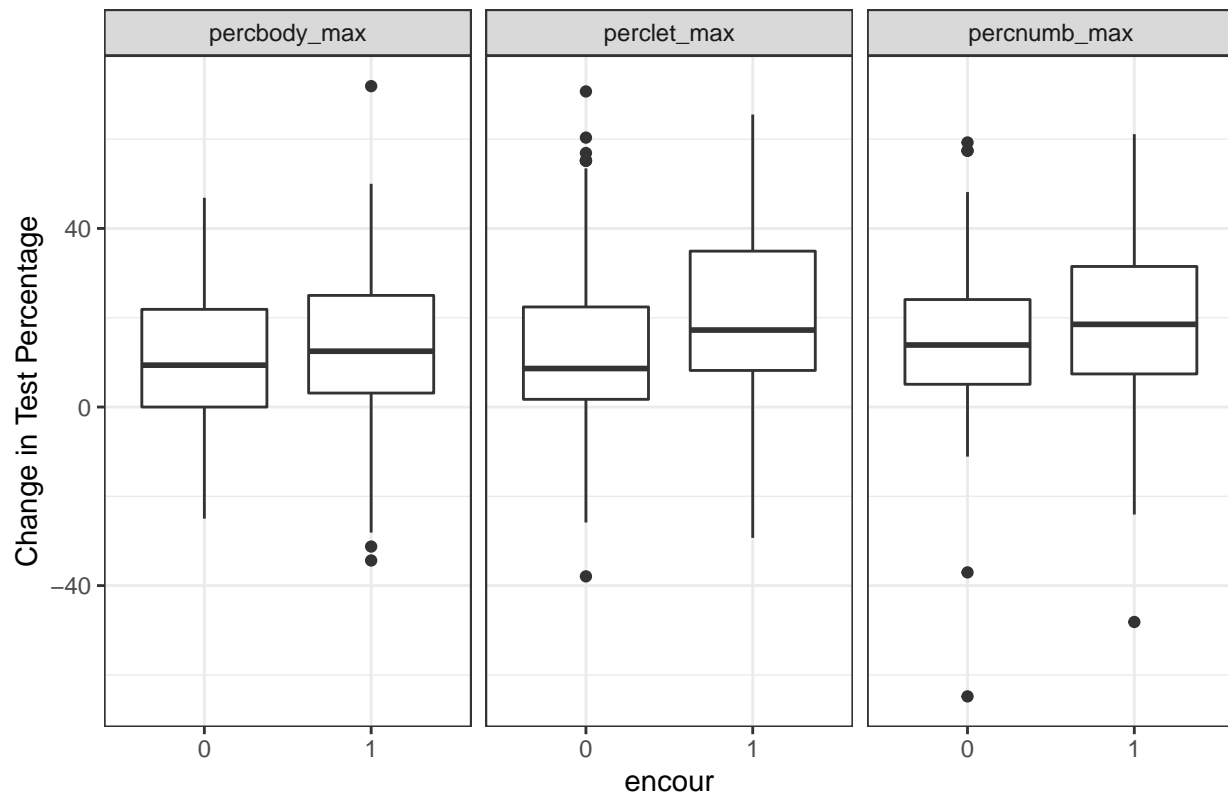
```
ggplot(ss_cat_long, aes(x = setting, y = value)) +
  geom_boxplot() +
  facet_wrap(~name) +
  labs(y = "Change in Test Percentage",
       title = "Child Change in Pre-Post Test Score Percentages vs Setting")
```

Child Change in Pre-Post Test Score Percentages vs Setting



```
ggplot(ss_cat_long, aes(x = encour, y = value)) +
  geom_boxplot() +
  facet_wrap(~name) +
  labs(y = "Change in Test Percentage",
       title = "Child Change in Pre-Post Test Score Percentages vs Encouragement")
```

Child Change in Pre-Post Test Score Percentages vs Encouragement



Correlation Plot

```
# correlation plots
ss_all_modified <- ss_all %>%
  select(percnumb_max, perclet_max, percbody_max, age, peabody)
ss_all_modified <- cor(ss_all_modified, use="pairwise.complete.obs")
corrplot(ss_all_modified, method = "number")
```

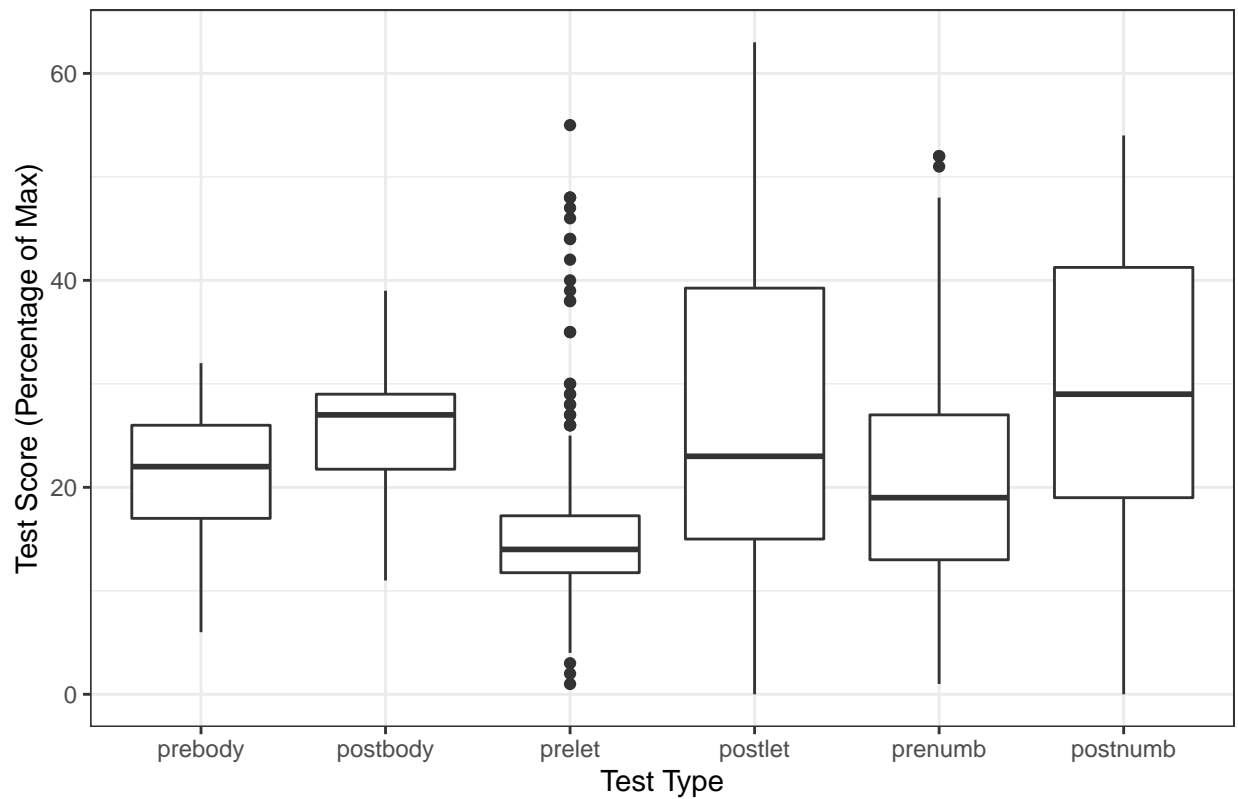



Test Score Distributions

```
# factor levels set to make sure boxplot is in the desired order
ss_all %>%
  select(id, prenumb, postnumb, prelet, postlet, prebody, postbody) %>%
  pivot_longer(-id) %>%
  mutate(name = factor(name, levels = c("prebody", "postbody", "prelet", "postlet",
    "prenumb", "postnumb"))) %>%

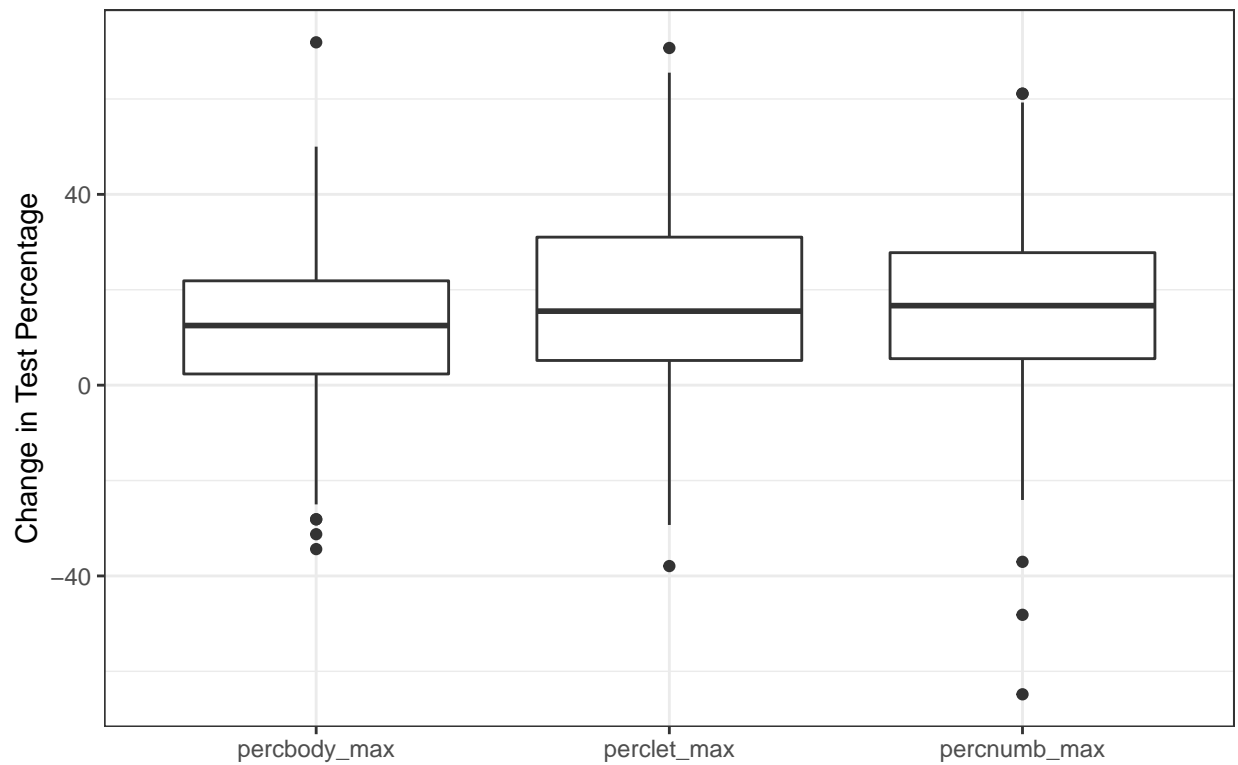
  ggplot(aes(x = name, y = value)) +
  geom_boxplot() +
  labs(y = "Test Score (Percentage of Max)", x = "Test Type",
    title = "Distribution of test scores for all pre/post tests")
```

Distribution of test scores for all pre/post tests



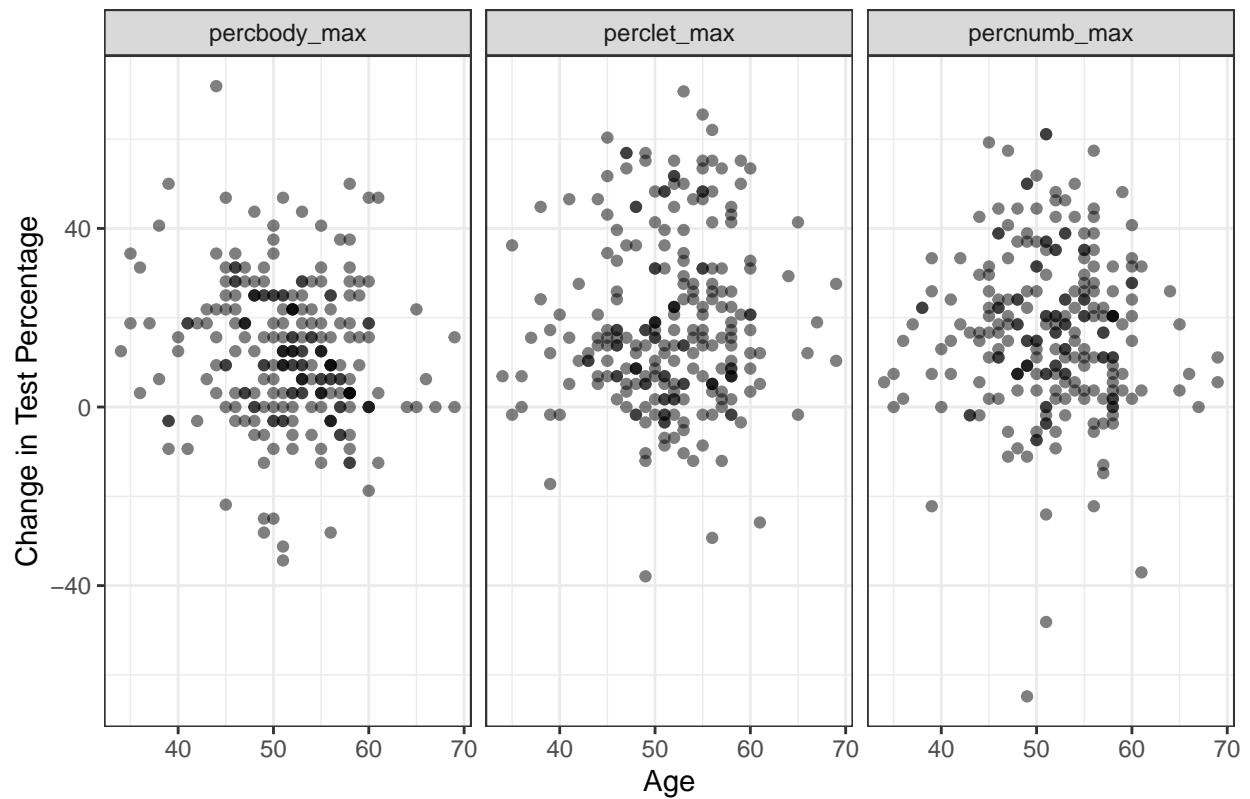
```
# Distributions of test score changes, split by category
ss_all %>%
  select(id, percnumb_max, perclet_max, percbbody_max) %>%
  pivot_longer(-id) %>%
  ggplot(aes(x = name, y = value)) +
  geom_boxplot() +
  labs(y = "Change in Test Percentage", x = "",
       title = "Child Change in Pre-Post Test Score Percentages")
```

Child Change in Pre-Post Test Score Percentages



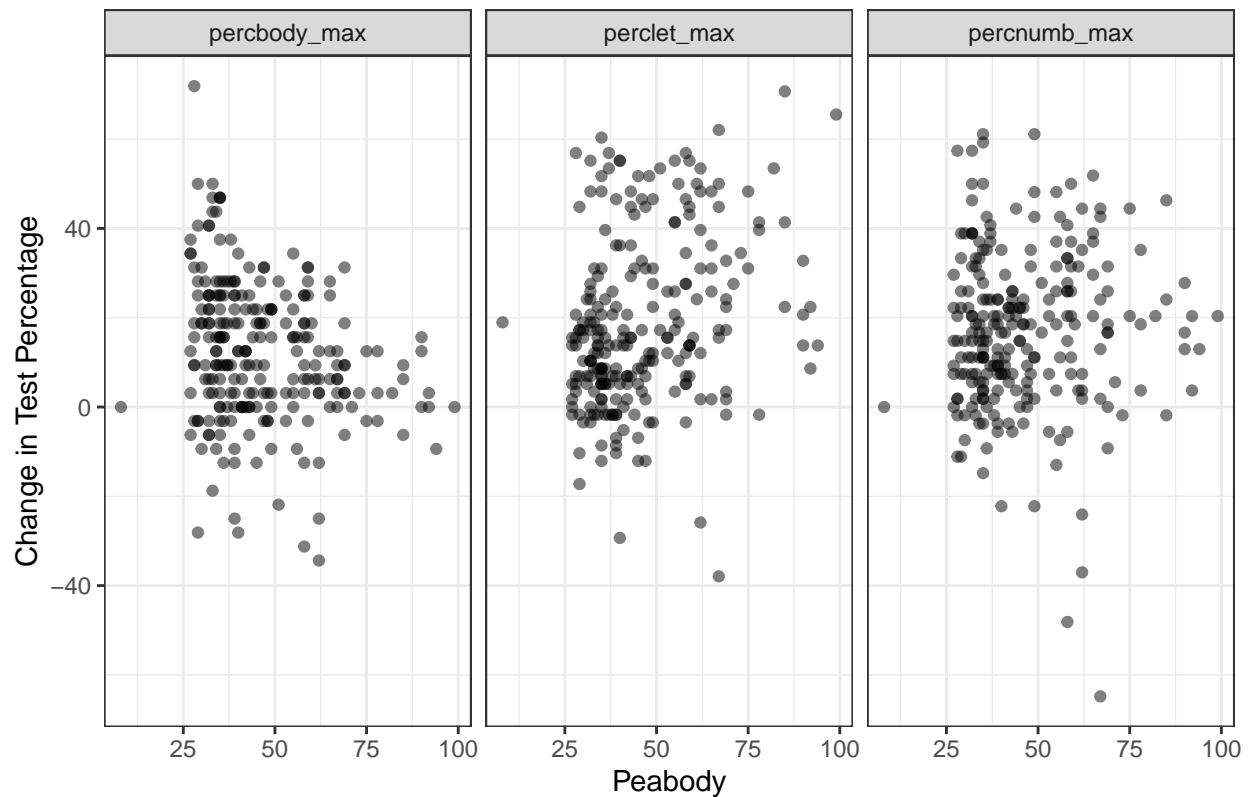
```
# Scatterplot of test score vs age, separated by test type
ss_all %>%
  select(id, percnumb_max, perclet_max, percbody_max, age, peabody) %>%
  pivot_longer(cols = c(percnumb_max, perclet_max, percbody_max)) %>%
  ggplot(aes(x = age, y = value)) +
  geom_point(alpha = 0.5) +
  facet_wrap(~name) +
  labs(y = "Change in Test Percentage", x = "Age",
       title = "Child Change in Pre-Post Test Score Percentages vs Age")
```

Child Change in Pre-Post Test Score Percentages vs Age



```
# Scatterplot of test score vs peabody, separated by test type
ss_all %>%
  select(id, percnumb_max, perclet_max, percbod_max, age, peabody) %>%
  pivot_longer(cols = c(percnumb_max, perclet_max, percbod_max)) %>%
  ggplot(aes(x = peabody, y = value)) +
  geom_point(alpha = 0.5) +
  facet_wrap(~name) +
  labs(y = "Change in Test Percentage", x = "Peabody",
       title = "Child Change in Pre-Post Test Score Percentages vs Peabody")
```

Child Change in Pre–Post Test Score Percentages vs Peabody



```
ss_all_dep <-
  ss_all %>%
  select(age, peabody, site)

groupnames <- colnames(ss_all_dep)

for (i in groupnames) {
  plot(ss_all_dep[[i]], ss_all$percnumb_max, ylab="percnum_max", xlab=i, main=paste("Scatterplot of percnumb_max vs", i))
}

for (i in groupnames) {
  plot(ss_all_dep[[i]], ss_all$perclet_max, ylab="perclet_max", xlab=i, main=paste("Scatterplot of perclet_max vs", i))
}

for (i in groupnames) {
  plot(ss_all_dep[[i]], ss_all$percbody_max, ylab="percbody_max", xlab=i, main=paste("Scatterplot of percbody_max vs", i))
}
```