DATAFEST!!!

.gif or .gif

2024-04-05

```
#Import data
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
## -- 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   1.5.0
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.1
## 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
```

```
page_views <- read.csv("~/Desktop/DataFest/page_views.csv")
#View(page_views)

responses <- read.csv("~/Desktop/DataFest/responses.csv", comment.char="#")#View(responses_sample)
```

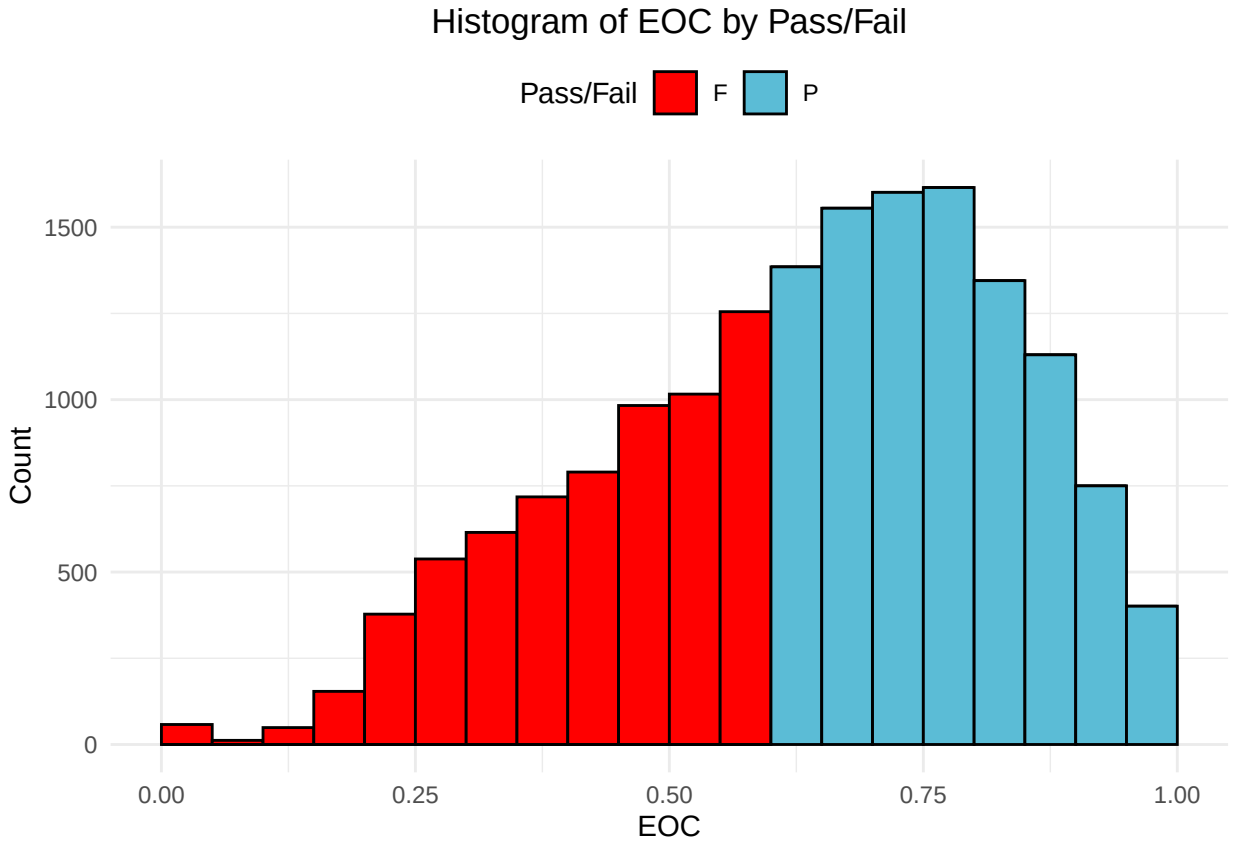
```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec,
## : EOF within quoted string
```

```
#View(responses)

media_views <- read.csv("~/Desktop/DataFest/media_views.csv")
#View(media_views)

items <- read.csv("~/Desktop/DataFest/items.csv")
#View(items)

checkpoints_pulse <- read.csv("~/Desktop/DataFest/checkpoints_pulse.csv")
#View(checkpoints_pulse)

checkpoints_eoc <- read.csv("~/Desktop/DataFest/checkpoints_eoc.csv")
#View(checkpoints_eoc)
```

```
#Creating success variable by .6 > is pass and below is fail
checkpoints_eoc$success <- ifelse(checkpoints_eoc$EOC > .6, "P", "F")

#Histogram of EOC
```

```r
#install.packages("wesanderson")
library(wesanderson)

#colors for plot
desired_color <- wes_palette("Darjeeling1")[1]
desired_color2 <- wes_palette("Darjeeling1")[5]


ggplot(data = checkpoints_eoc, aes(x = EOC, fill = success)) +
  geom_histogram(color = "black", binwidth = 0.1, breaks = c(0, 0.05, 0.1, .15, 0.2, .25, 0.3, .35, 0.4
  labs(x = "EOC",
       y = "Count",
       title = "Histogram of EOC by Pass/Fail",
       fill = "Pass/Fail") +
  scale_fill_manual(values = c(desired_color, desired_color2)) +
  theme_minimal() +
  theme(legend.position = "top", plot.title = element_text(hjust = 0.5))
```

```
## Warning: Removed 70 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

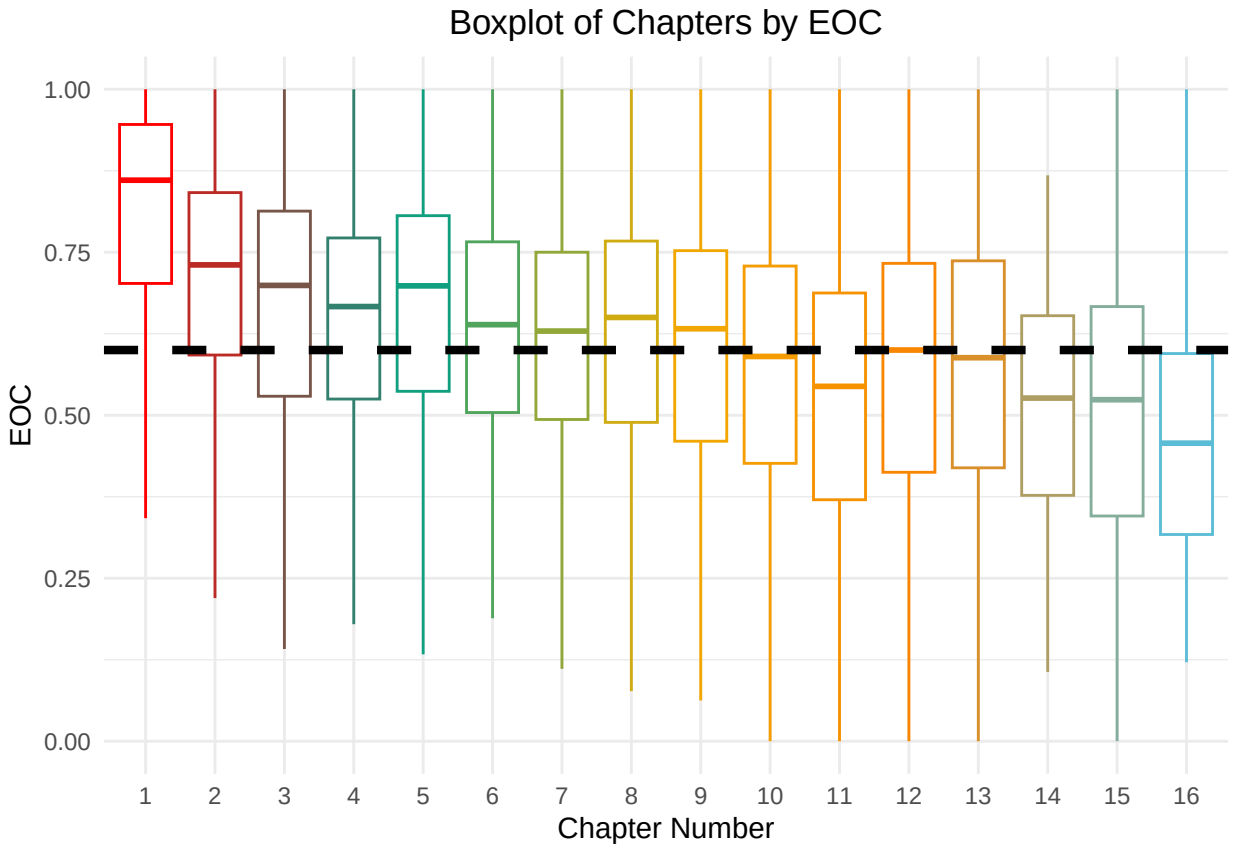



```r
#Boxplot of chapters by EOC
ggplot(data = checkpoints_eoc, aes(x = as.factor(chapter_number), y = EOC)) +
  geom_boxplot(outliers = FALSE, aes(color = as.factor(chapter_number)), show.legend = FALSE) +
  labs(title = "Boxplot of Chapters by EOC",
       x = "Chapter Number") +
  scale_color_manual(values = wes_palette("Darjeeling1", n = 16, type = "continuous")) +
```

```
  theme_minimal() +
  geom_hline(yintercept = 0.6, linetype = "dashed", color = "black", size = 1.5) +
  theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

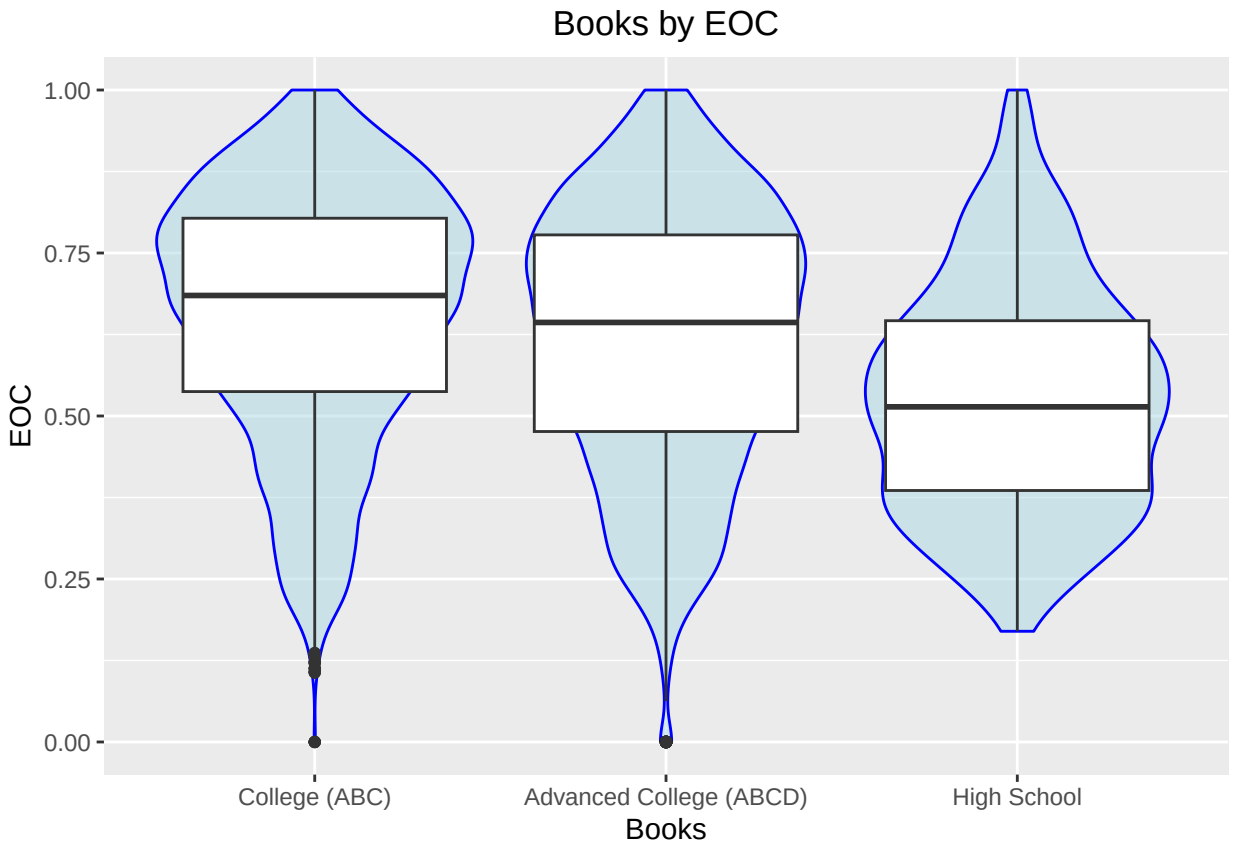
```
## Warning: Removed 70 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
#Boxplot and violin plot of books by EOC
ggplot(checkpoints_eoc, aes(x = book, y = EOC)) +
  geom_violin(fill = "lightblue", color = "blue", alpha = 0.5) +
  geom_boxplot() +
  labs(x = "Books",
       title = "Books by EOC") +
  scale_x_discrete(labels = c("College (ABC)", "Advanced College (ABCD)", "High School")) +
  theme(plot.title = element_text(hjust = 0.5))
```

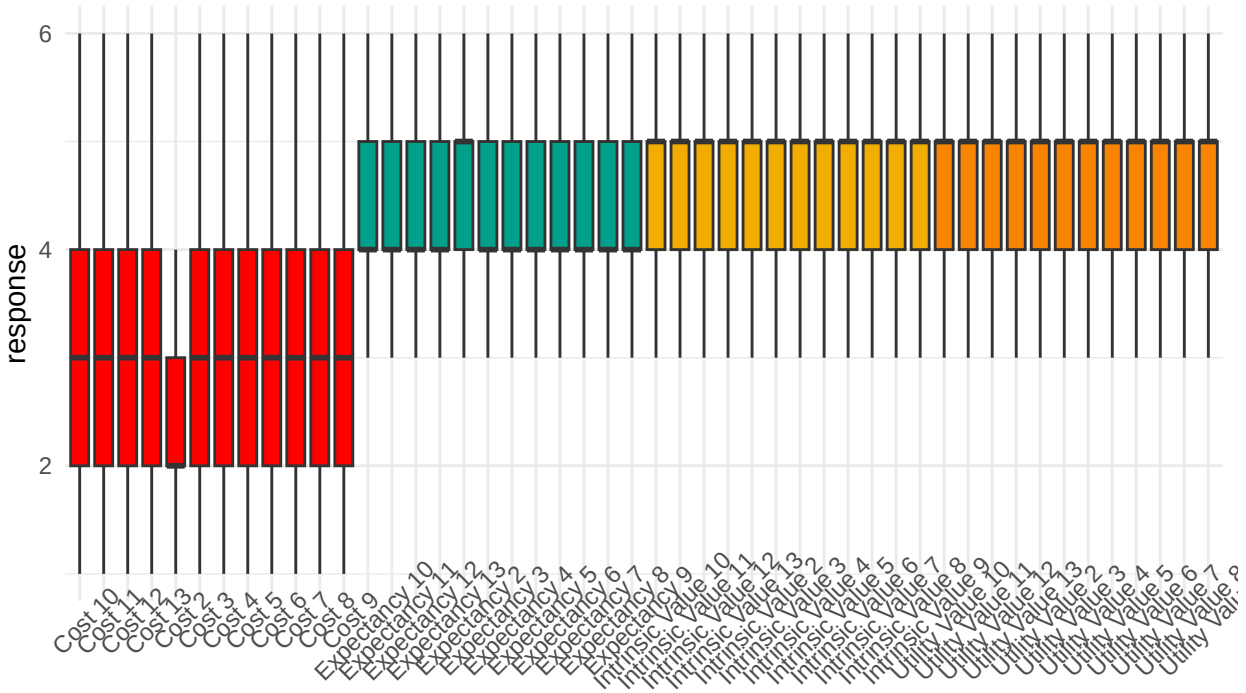
```
## Warning: Removed 70 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 70 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
#Boxplot of created combo variable with a fill of construct
check <- checkpoints_pulse %>% mutate(combo = paste(construct, chapter_number, sep = " "))

ggplot(check, aes(x=combo,y=response, fill = construct)) +
  geom_boxplot(outliers = F) +
  theme_minimal() +
  labs(title = "Boxplot of Pulse Responses by Construct and Chapter") +
  scale_fill_manual(values = wes_palette(name="Darjeeling1",n=4,type = "discrete")) +
  theme(legend.position = "none",axis.text.x=element_text(angle=45),plot.title = element_text(hjust = 0
```

```
## Warning: Removed 32386 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
filtered_eoc <- checkpoints_eoc %>%
  mutate(avg_attempt = n_attempt/n_possible) %>% #creating metric to see how many attempts per possible
  filter(book == "College / Statistics and Data Science (ABC)", !is.na(EOC)) %>% #filtering by book wit
  select(-c(n_possible,n_correct,n_attempt,book))

filtered_views <- page_views %>%
  filter(book == "College / Statistics and Data Science (ABC)") %>%
  mutate(idle = idle_brief + idle_long) %>% #combining both similar time columns
  mutate(off_page = off_page_brief + off_page_long) %>% #combining both similar time columns
  select(student_id, chapter_number, institution_id, release, engaged, idle, off_page, tried_again_click
  group_by(student_id, institution_id, chapter_number, release) %>%
  summarise(engaged_sum = sum(engaged, na.rm = T) / 60000, #convert to minutes from milliseconds
            idle_sum = sum(idle, na.rm = T) / 60000, #convert to minutes from milliseconds
            off_page_sum = as.numeric(format(sum(off_page, na.rm = T) / 60000, scientific = F)), #conve
            tried_again_clicks_sum = sum(tried_again_clicks, na.rm = T))
```

```
## `summarise()` has grouped output by 'student_id', 'institution_id',
## 'chapter_number'. You can override using the `.groups` argument.
```

```
DATA <- left_join(filtered_eoc,filtered_views, by = c("student_id","chapter_number")) # final table
```

```
## Warning in left_join(filtered_eoc, filtered_views, by = c("student_id", : Detected an unexpected many
## i Row 127 of `x` matches multiple rows in `y`.
## i Row 138 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
total_table <- left_join(filtered_eoc,filtered_views, by = c("student_id","chapter_number")) # final ta
```

```
## Warning in left_join(filtered_eoc, filtered_views, by = c("student_id", : Detected an unexpected many
## i Row 127 of `x` matches multiple rows in `y`.
## i Row 138 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

#code for all the pulse graphs

```
pulse <- checkpoints_pulse %>%
  mutate(cost = ifelse(construct == "Cost", response, NA)) %>%
  mutate(expectancy = ifelse(construct == "Expectancy", response, NA)) %>%
  mutate(intrinsic = ifelse(construct == "Intrinsic Value", response, NA)) %>%
  mutate(utility = ifelse(construct == "Utility Value", response, NA)) %>%
  filter(response != "")

pulse <- pulse %>% group_by(student_id, chapter_number) %>%
  summarize(cost = mean(cost, na.rm = TRUE),
            expectancy = mean(expectancy, na.rm = TRUE),
            intrinsic = mean(intrinsic, na.rm = TRUE),
            utility = mean(utility, na.rm = TRUE)) %>%
  filter(cost != 3.5,expectancy != 3.5)
```

```
## `summarise()` has grouped output by 'student_id'. You can override using the
## `.groups` argument.
```

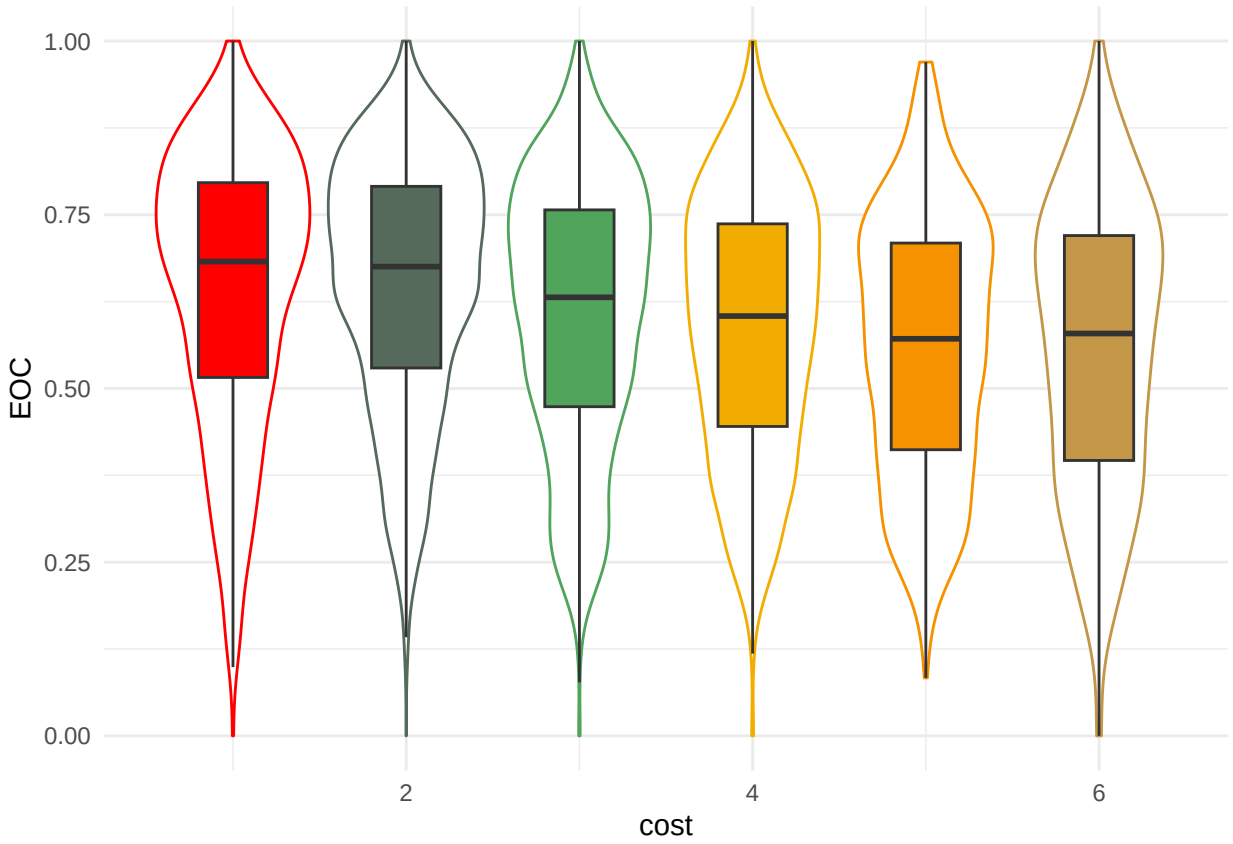
```
full_pulse <- full_join(total_table,pulse, by = c("student_id","chapter_number"))
```

```
ggplot(full_pulse, aes(x = cost, y = EOC)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(cost))) +
  geom_boxplot(aes(fill = as.factor(cost)), width = 0.4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  scale_color_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  theme(legend.position = "none")
```

```
## Warning: Removed 4870 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 1813 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 3057 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
eoc_cost <- lm(EOC ~ cost, full_pulse)
summary(eoc_cost)
```

```
##
## Call:
## lm(formula = EOC ~ cost, data = full_pulse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.66104 -0.13379  0.02359  0.14846  0.45963
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.685177   0.005533  123.84   <2e-16 ***
## cost        -0.024134   0.001730  -13.95   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1878 on 7918 degrees of freedom
##   (4870 observations deleted due to missingness)
## Multiple R-squared:  0.02398,    Adjusted R-squared:  0.02386
## F-statistic: 194.5 on 1 and 7918 DF,  p-value: < 2.2e-16
```

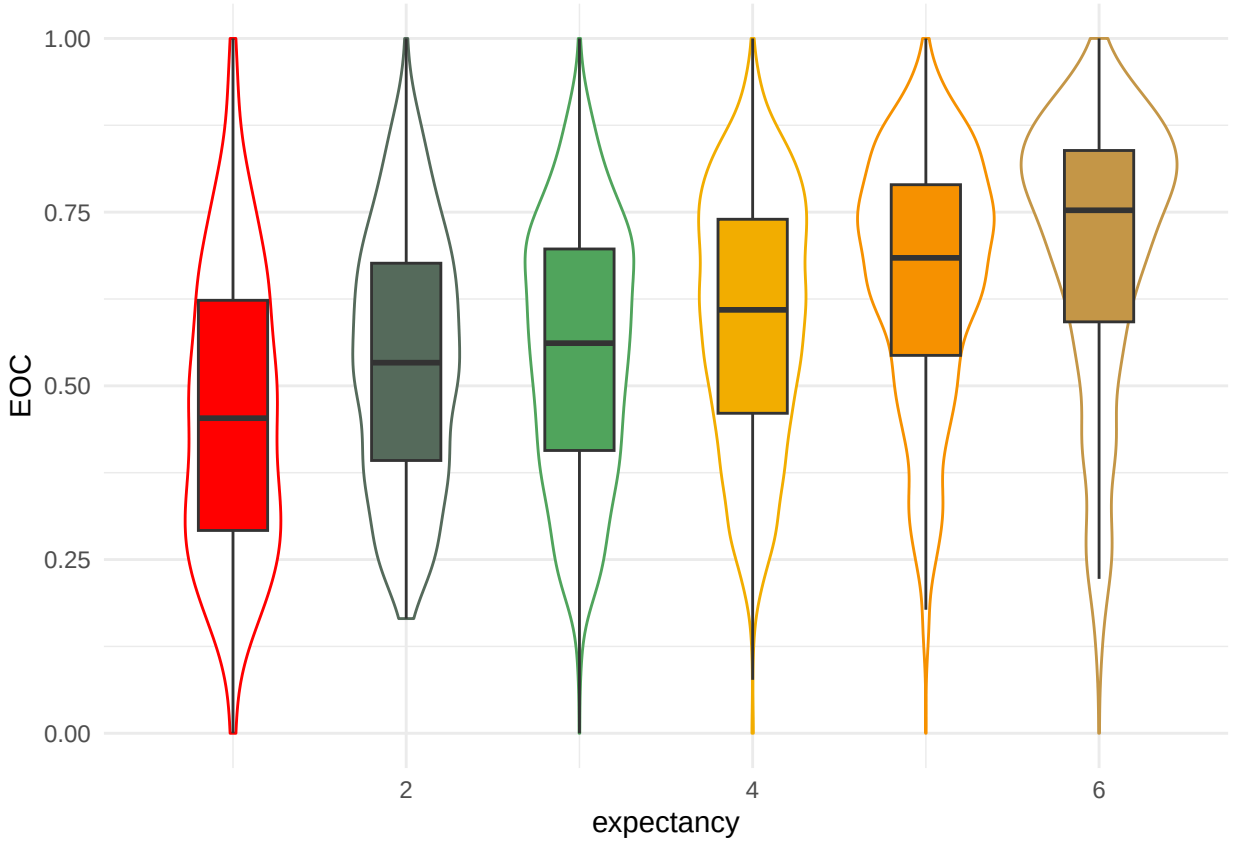
```
ggplot(full_pulse, aes(x = expectancy, y = EOC)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(expectancy))) +
  geom_boxplot(aes(fill = as.factor(expectancy)), width = 0.4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
```

```
  scale_color_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  theme(legend.position = "none")
```

```
## Warning: Removed 4870 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 1813 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 3057 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
eoc_exp <- lm(EOC ~ expectancy, full_pulse)
summary(eoc_exp)
```

```
##
## Call:
## lm(formula = EOC ~ expectancy, data = full_pulse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.69508 -0.12633  0.02348  0.14183  0.53882
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.414394   0.008495   48.78   <2e-16 ***
## expectancy  0.046782   0.001933   24.20   <2e-16 ***
## ---
```

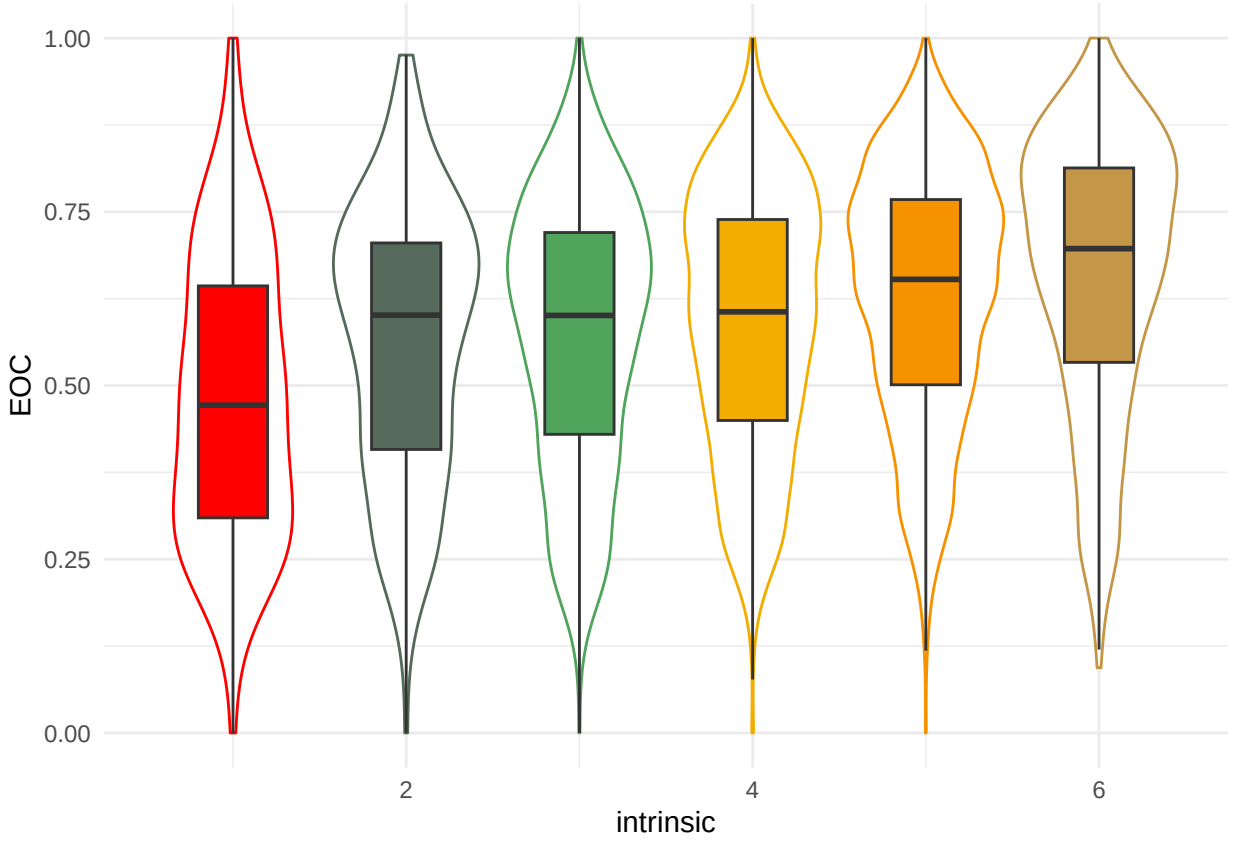
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1835 on 7918 degrees of freedom
##   (4870 observations deleted due to missingness)
## Multiple R-squared:  0.06888,    Adjusted R-squared:  0.06876
## F-statistic: 585.7 on 1 and 7918 DF,  p-value: < 2.2e-16
```

```
ggplot(full_pulse, aes(x = intrinsic, y = EOC)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(intrinsic))) +
  geom_boxplot(aes(fill = as.factor(intrinsic)), width = 0.4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  scale_color_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  theme(legend.position = "none")
```

```
## Warning: Removed 5153 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 2187 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 2966 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
eoc_int <- lm(EOC ~ intrinsic, full_pulse)
summary(eoc_int)
```

```
##
## Call:
```

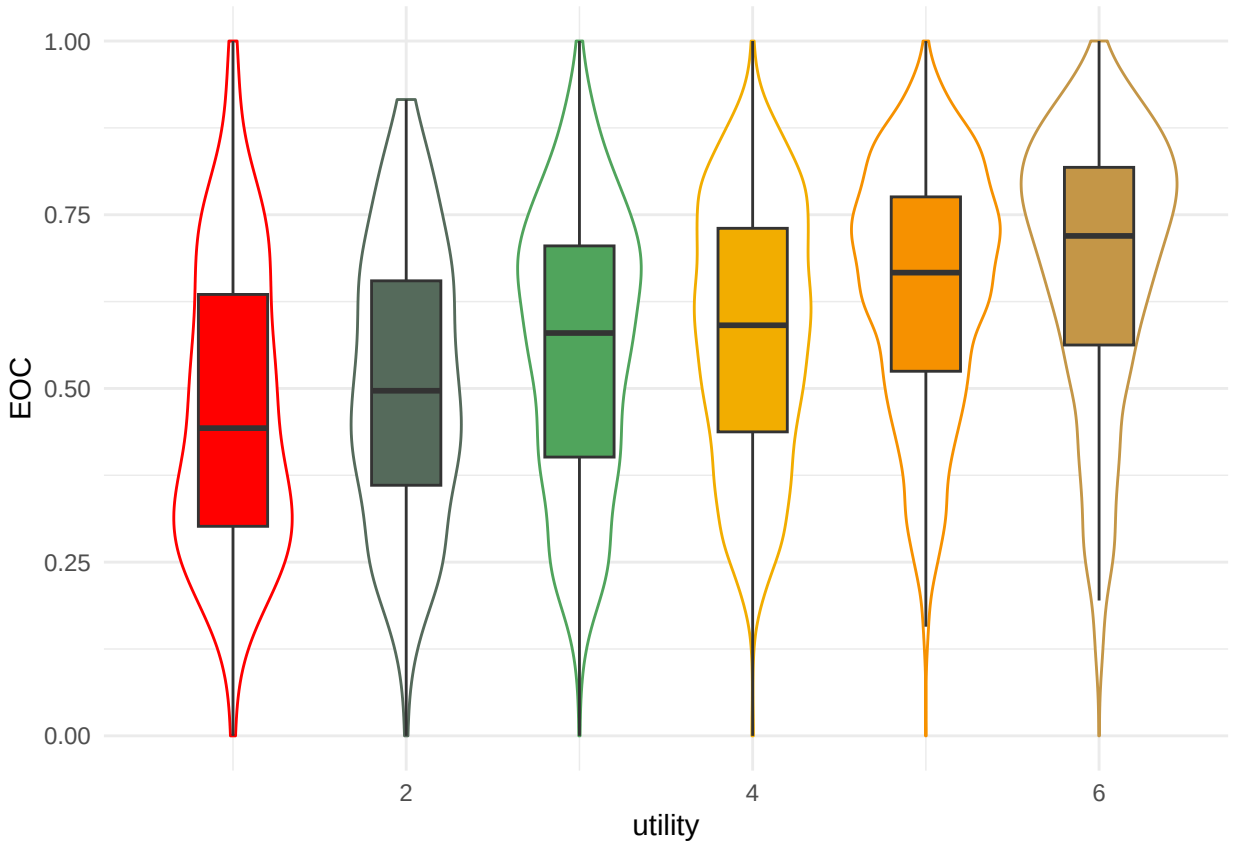
```
## lm(formula = EOC ~ intrinsic, data = full_pulse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62761 -0.13339  0.02316  0.14605  0.49928
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.468999   0.009838   47.67   <2e-16 ***
## intrinsic   0.031721   0.002122   14.95   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1874 on 7635 degrees of freedom
##   (5153 observations deleted due to missingness)
## Multiple R-squared:  0.02843,    Adjusted R-squared:  0.0283
## F-statistic: 223.4 on 1 and 7635 DF,  p-value: < 2.2e-16
```

```
ggplot(full_pulse, aes(x = utility, y = EOC)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(utility))) +
  geom_boxplot(aes(fill = as.factor(utility)), width = 0.4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  scale_color_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  theme(legend.position = "none")
```

```
## Warning: Removed 4948 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 1930 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 3018 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
eoc_util <- lm(EOC ~ utility, full_pulse)
summary(eoc_util)
```

```
##
## Call:
## lm(formula = EOC ~ utility, data = full_pulse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.67916 -0.13088  0.02291  0.14275  0.54392
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.411459   0.009913   41.51   <2e-16 ***
## utility     0.044617   0.002135   20.90   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1849 on 7840 degrees of freedom
##   (4948 observations deleted due to missingness)
## Multiple R-squared:  0.05276,    Adjusted R-squared:  0.05264
## F-statistic: 436.7 on 1 and 7840 DF,  p-value: < 2.2e-16
```

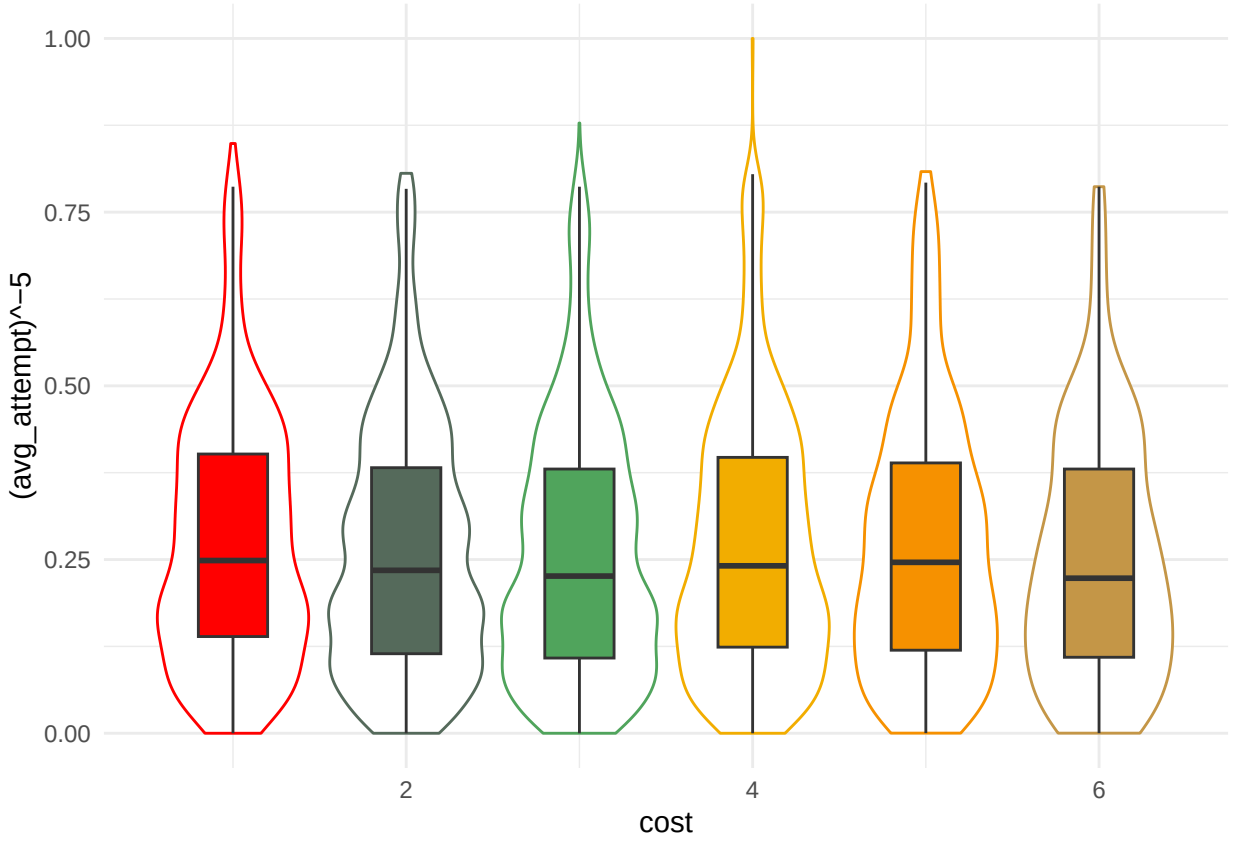
```
ggplot(full_pulse, aes(y = (avg_attempt)^-5, x = cost)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(cost))) +
  geom_boxplot(aes(fill = as.factor(cost)), width = 0.4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
```

```
  scale_color_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  theme(legend.position = "none")
```

```
## Warning: Removed 4870 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 1813 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 3057 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
att_cost <- lm((avg_attempt)^-5 ~ cost, full_pulse)
summary(att_cost)
```

```
##
## Call:
## lm(formula = (avg_attempt)^-5 ~ cost, data = full_pulse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2672 -0.1500 -0.0297  0.1220  0.7331
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.2673155  0.0054064  49.445   <2e-16 ***
## cost        -0.0001044  0.0016907  -0.062    0.951
## ---
```

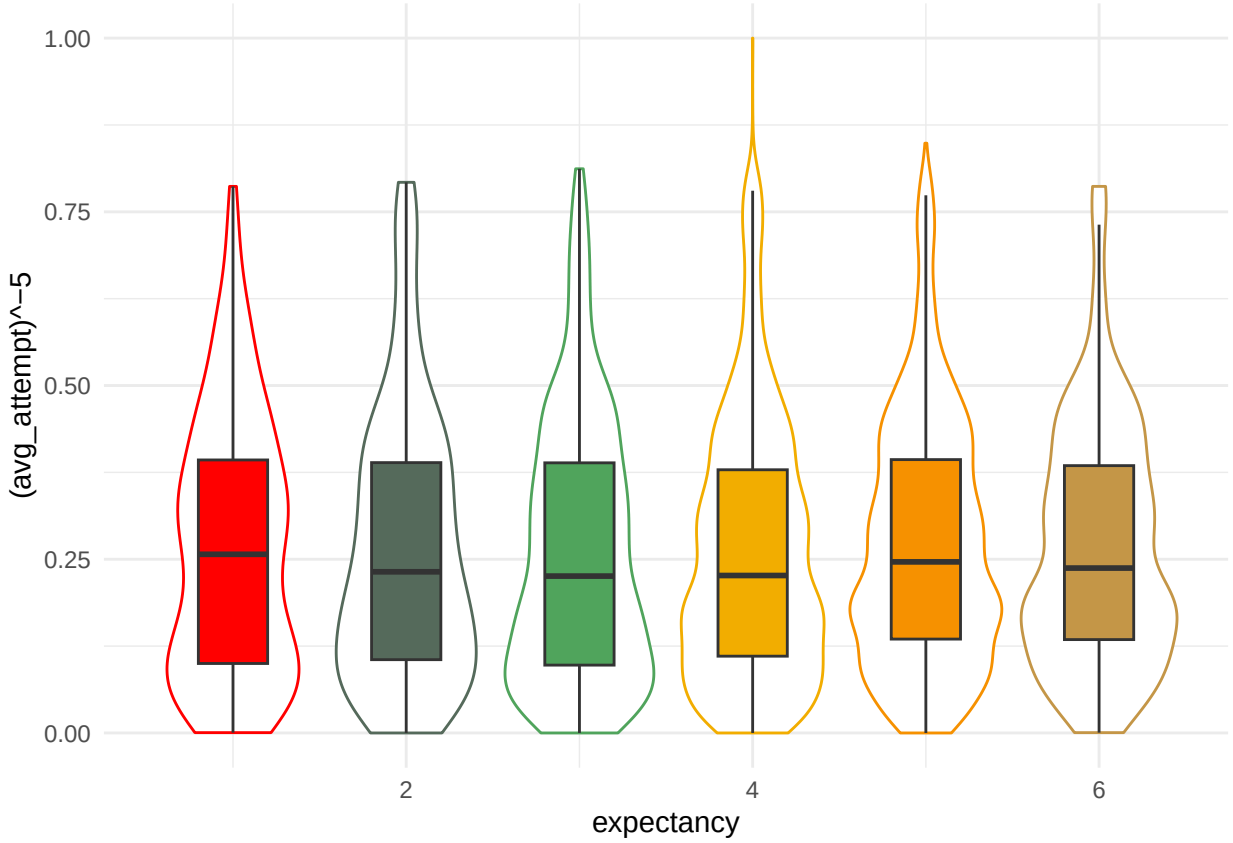
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1835 on 7918 degrees of freedom
##   (4870 observations deleted due to missingness)
## Multiple R-squared:  4.819e-07,  Adjusted R-squared:  -0.0001258
## F-statistic: 0.003816 on 1 and 7918 DF,  p-value: 0.9507
```

```
ggplot(full_pulse, aes(y = (avg_attempt)^-5, x = expectancy)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(expectancy))) +
  geom_boxplot(aes(fill = as.factor(expectancy)), width = 0.4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  scale_color_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  theme(legend.position = "none")
```

```
## Warning: Removed 4870 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 1813 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 3057 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
att_exp <- lm((avg_attempt)^-5 ~ expectancy, full_pulse)
summary(att_exp)
```

```
##
## Call:
```

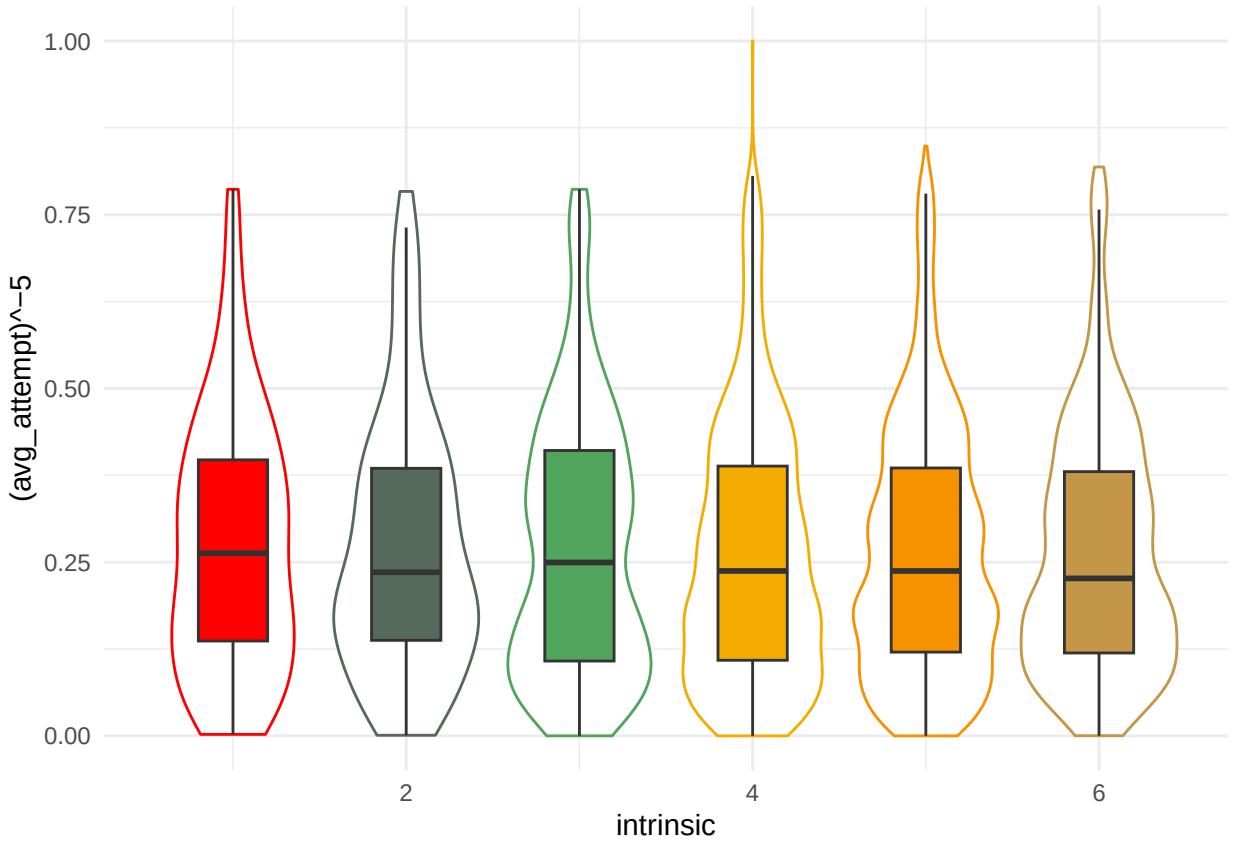
```
## lm(formula = (avg_attempt)^-5 ~ expectancy, data = full_pulse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27187 -0.14970 -0.02889  0.12106  0.73380
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.253912   0.008497  29.882   <2e-16 ***
## expectancy  0.003072   0.001933   1.589    0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1835 on 7918 degrees of freedom
##   (4870 observations deleted due to missingness)
## Multiple R-squared:  0.0003186,  Adjusted R-squared:  0.0001924
## F-statistic: 2.524 on 1 and 7918 DF,  p-value: 0.1122
```

```
ggplot(full_pulse, aes(y = (avg_attempt)^-5, x = intrinsic)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(intrinsic))) +
  geom_boxplot(aes(fill = as.factor(intrinsic)), width = 0.4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  scale_color_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  theme(legend.position = "none")
```

```
## Warning: Removed 5153 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 2187 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 2966 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
att_int <- lm((avg_attempt)^-5 ~ intrinsic, full_pulse)
summary(att_int)
```

```
##
## Call:
## lm(formula = (avg_attempt)^-5 ~ intrinsic, data = full_pulse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2702 -0.1511 -0.0304  0.1220  0.7323
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.273789   0.009658   28.35   <2e-16 ***
## intrinsic   -0.001520   0.002084   -0.73    0.466
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.184 on 7635 degrees of freedom
##   (5153 observations deleted due to missingness)
## Multiple R-squared:  6.971e-05,  Adjusted R-squared:  -6.126e-05
## F-statistic: 0.5323 on 1 and 7635 DF,  p-value: 0.4657
```

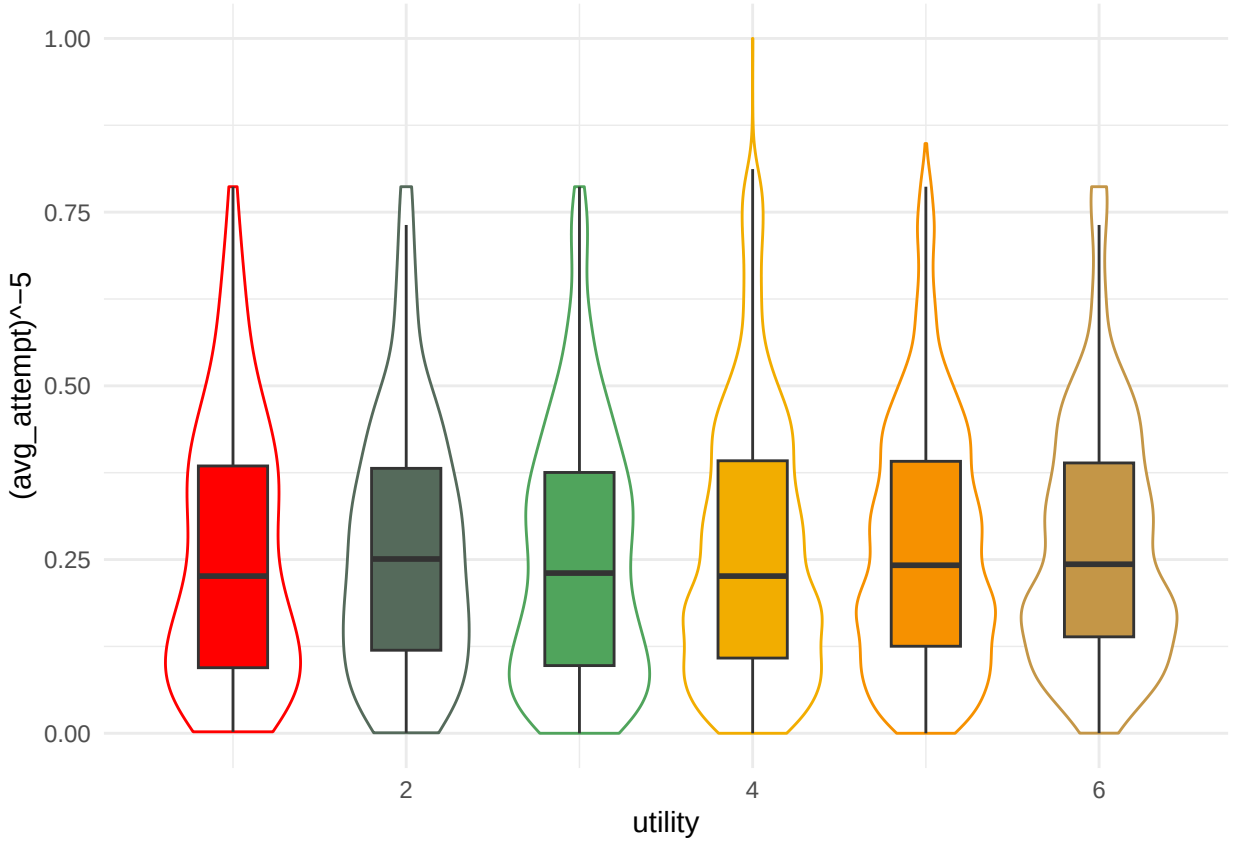
```
ggplot(full_pulse, aes(y = (avg_attempt)^-5, x = utility)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(utility))) +
  geom_boxplot(aes(fill = as.factor(utility)), width = 0.4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
```

```
  scale_color_manual(values = wes_palette(name = "Darjeeling1",n=7, type = "continuous")) +
  theme(legend.position = "none")
```

```
## Warning: Removed 4948 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 1930 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 3018 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
att_util <- lm((avg_attempt)^-5 ~ utility, full_pulse)
summary(att_util)
```

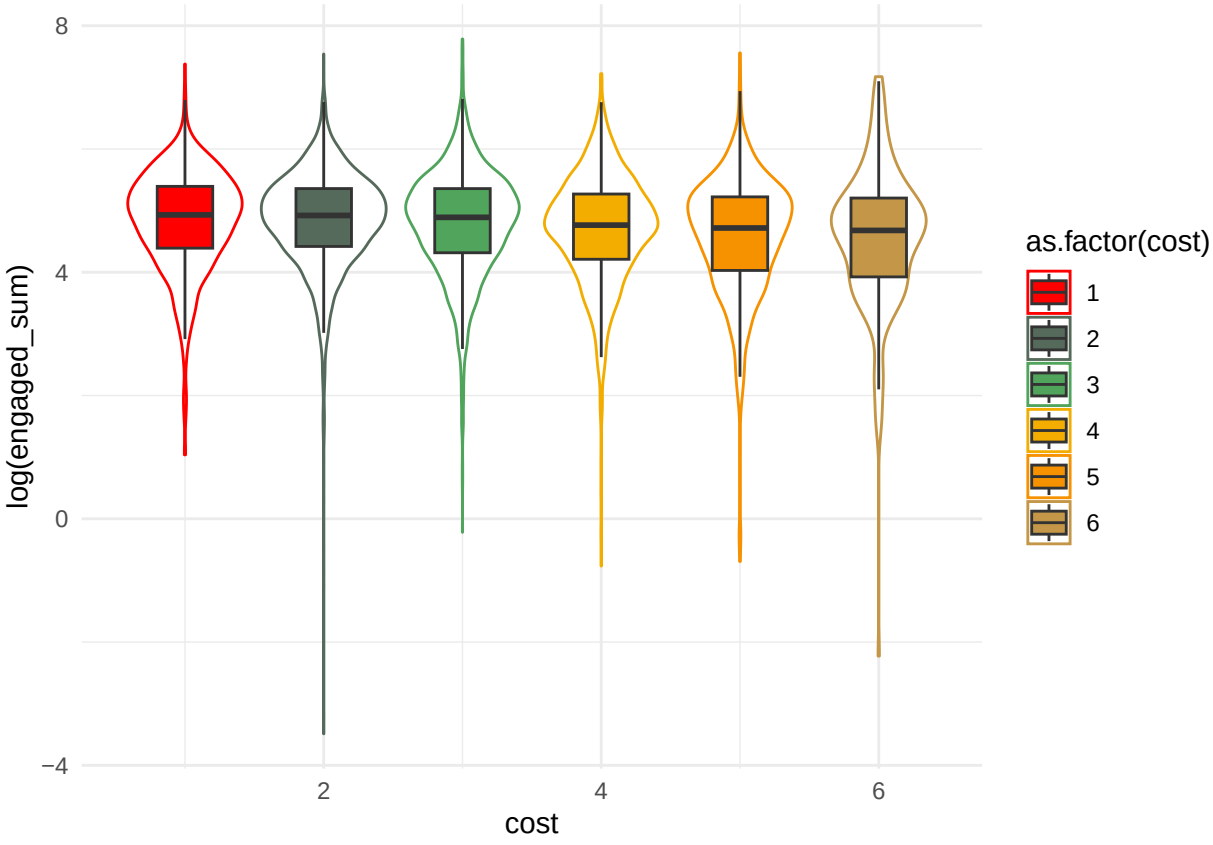
```
##
## Call:
## lm(formula = (avg_attempt)^-5 ~ utility, data = full_pulse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27257 -0.15009 -0.03074  0.12063  0.73487
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.249857   0.009844  25.381   <2e-16 ***
## utility     0.003819   0.002120   1.801   0.0717 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1837 on 7840 degrees of freedom
##   (4948 observations deleted due to missingness)
## Multiple R-squared:  0.0004137,  Adjusted R-squared:  0.0002862
## F-statistic: 3.245 on 1 and 7840 DF,  p-value: 0.0717
```

```
full_pulse <- full_pulse %>%
  filter(cost != 3.5)
ggplot(full_pulse, aes(y = log(engaged_sum), x = cost)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(cost))) +
  geom_boxplot(aes(fill = as.factor(cost)), width = .4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1", n = 7, type = "continuous")) +
  scale_color_manual(values = wes_palette(name = "Darjeeling1", n = 7, type = "continuous"))
```

```
## Warning: Removed 3062 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 3062 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



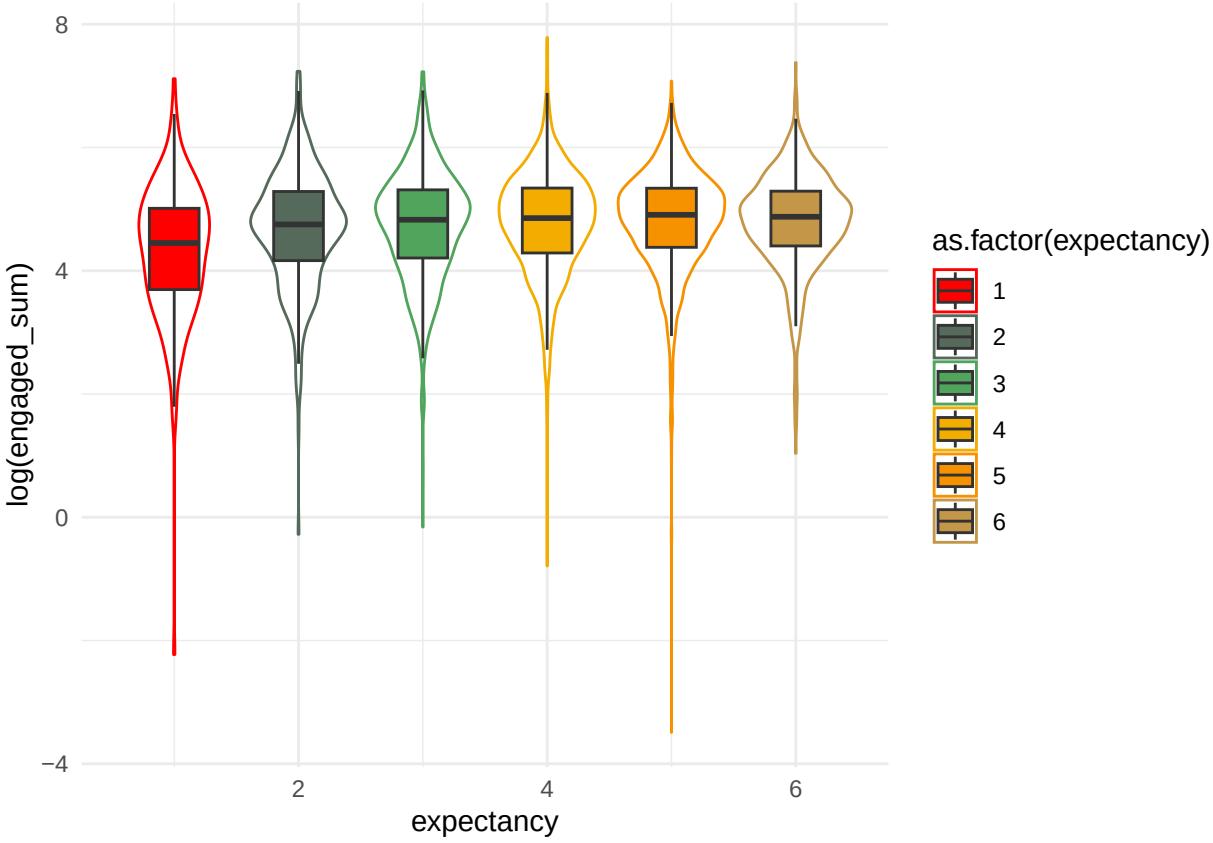

```
eng_cost <- lm(engaged_sum ~ cost, full_pulse)
summary(eng_cost)
```

```
##
## Call:
## lm(formula = engaged_sum ~ cost, data = full_pulse)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -169.84  -87.41  -32.67   43.91 2239.99
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  173.966      4.265  40.792  < 2e-16 ***
## cost          -4.128      1.334  -3.095  0.00197 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 144.8 on 7918 degrees of freedom
##   (3057 observations deleted due to missingness)
## Multiple R-squared:  0.001209,   Adjusted R-squared:  0.001082
## F-statistic: 9.581 on 1 and 7918 DF,  p-value: 0.001973
```

```
ggplot(full_pulse, aes(y = log(engaged_sum), x = expectancy)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(expectancy))) +
  geom_boxplot(aes(fill = as.factor(expectancy)), width = .4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1", n = 7, type = "continuous")) +
  scale_color_manual(values = wes_palette(name = "Darjeeling1", n = 7, type = "continuous"))
```

```
## Warning: Removed 3062 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
## Removed 3062 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



```r
eng_exp <- lm(engaged_sum ~ expectancy, full_pulse)
summary(eng_exp)
```

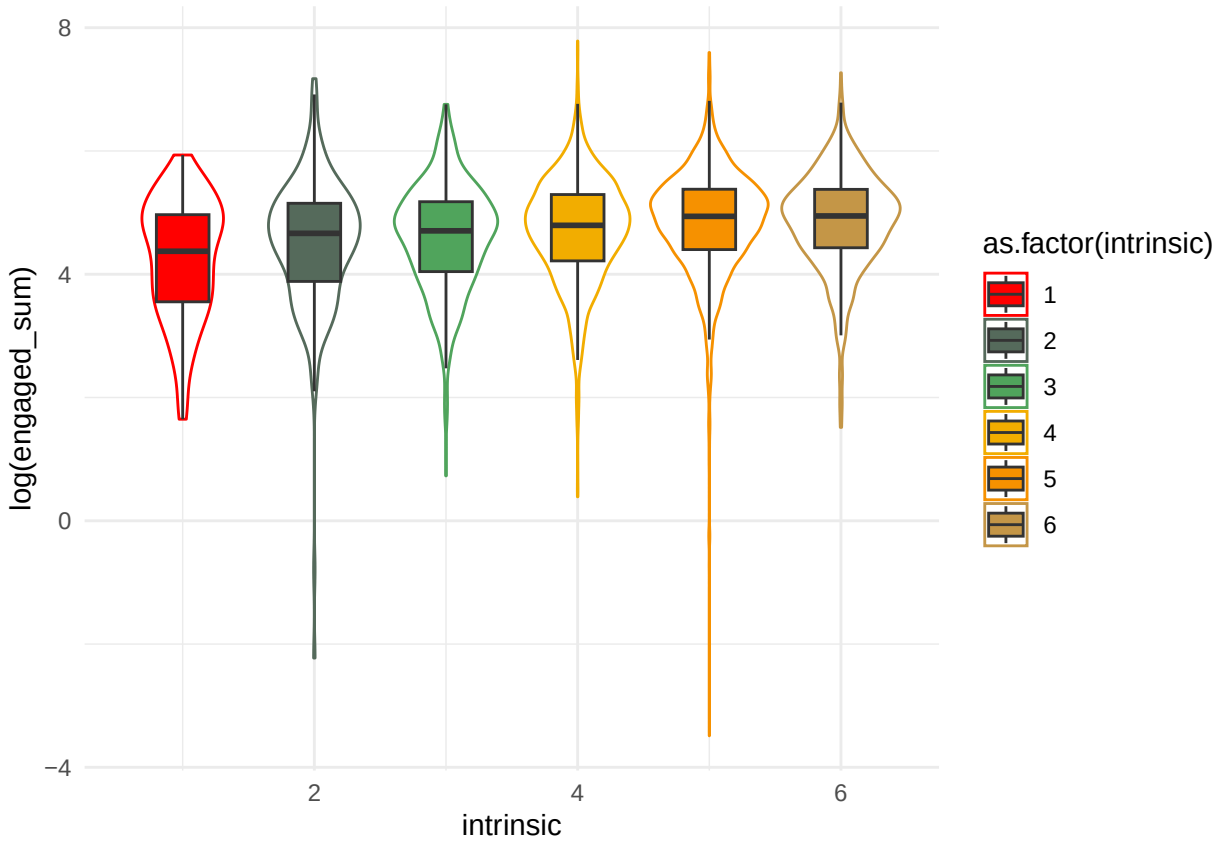
```
##
## Call:
## lm(formula = engaged_sum ~ expectancy, data = full_pulse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -164.19  -87.39  -32.64   43.24 2240.18
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  155.816      6.708  23.230   <2e-16 ***
## expectancy     1.395      1.526   0.914    0.361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 144.9 on 7918 degrees of freedom
##   (3057 observations deleted due to missingness)
## Multiple R-squared:  0.0001055,  Adjusted R-squared:  -2.077e-05
## F-statistic: 0.8355 on 1 and 7918 DF,  p-value: 0.3607
```

```r
ggplot(full_pulse, aes(y = log(engaged_sum), x = intrinsic)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(intrinsic))) +
  geom_boxplot(aes(fill = as.factor(intrinsic)), width = .4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1", n = 7, type = "continuous")) +
  scale_color_manual(values = wes_palette(name = "Darjeeling1", n = 7, type = "continuous"))
```

```
## Warning: Removed 3345 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 374 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 2971 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
eng_int <- lm(engaged_sum ~ intrinsic, full_pulse)
summary(eng_int)
```

```
##
## Call:
## lm(formula = engaged_sum ~ intrinsic, data = full_pulse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -179.65  -88.02  -32.40   43.62 2246.35
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  106.368      7.611  13.975  < 2e-16 ***
## intrinsic     12.213      1.642   7.438 1.13e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 145 on 7635 degrees of freedom
##   (3340 observations deleted due to missingness)
## Multiple R-squared:  0.007195,   Adjusted R-squared:  0.007065
## F-statistic: 55.33 on 1 and 7635 DF,  p-value: 1.13e-13
```

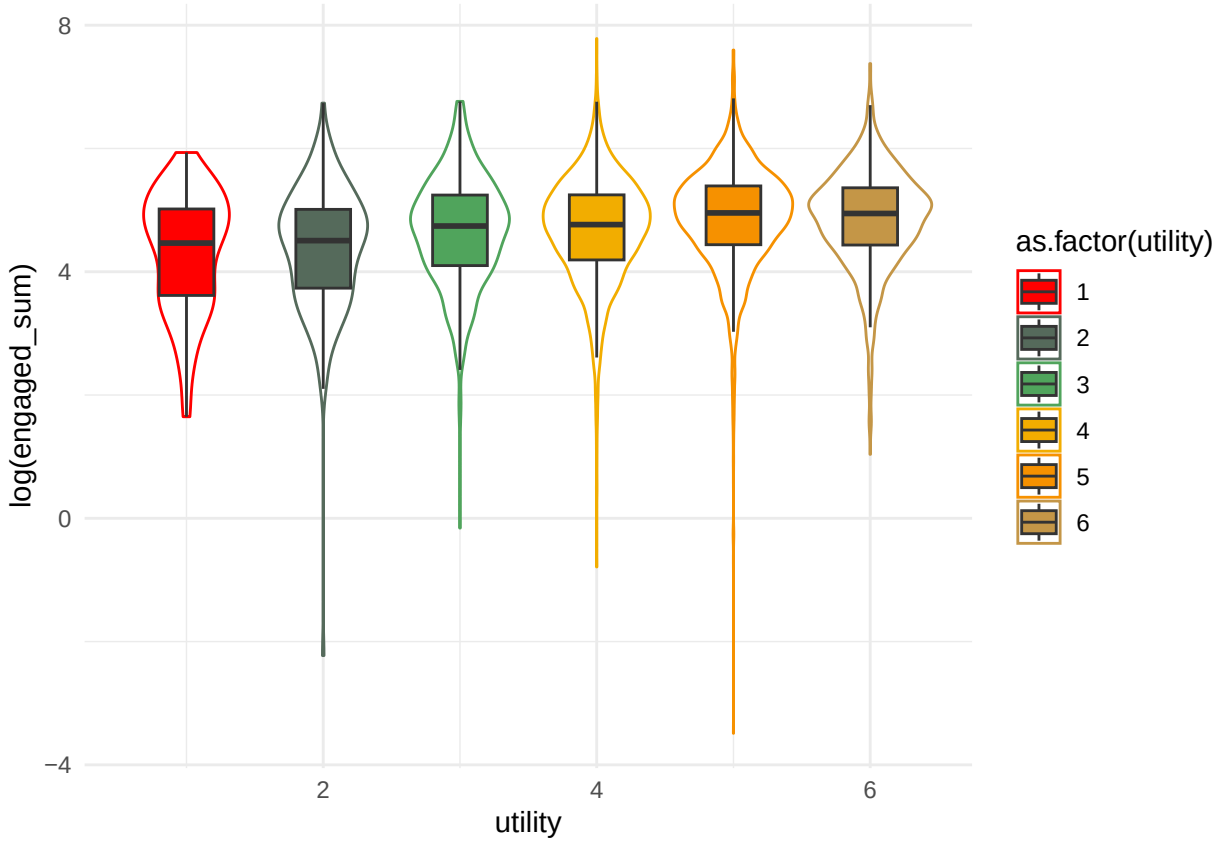
```
ggplot(full_pulse, aes(y = log(engaged_sum), x = utility)) +
  theme_minimal() +
  geom_violin(aes(color = as.factor(utility))) +
  geom_boxplot(aes(fill = as.factor(utility)), width = .4, outliers = F) +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1", n = 7, type = "continuous")) +
```

```
  scale_color_manual(values = wes_palette(name = "Darjeeling1", n = 7, type = "continuous"))
```

```
## Warning: Removed 3140 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

```
## Warning: Removed 117 rows containing missing values or values outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 3023 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
eng_util <- lm(engaged_sum ~ utility, full_pulse)
summary(eng_util)
```

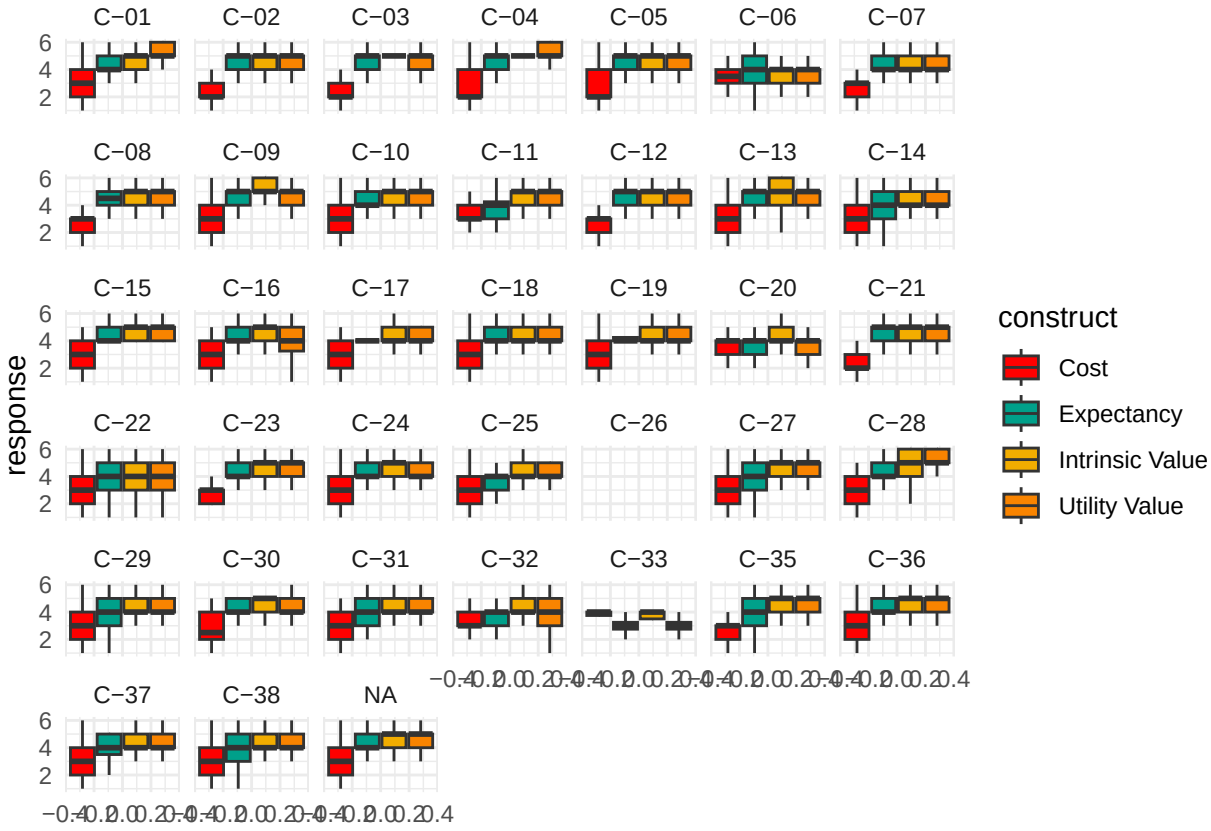
```
##
## Call:
## lm(formula = engaged_sum ~ utility, data = full_pulse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -181.40  -87.29  -32.10   43.85 2246.68
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  101.889      7.755  13.138  < 2e-16 ***
## utility       13.252      1.670   7.933 2.43e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 144.7 on 7840 degrees of freedom
##   (3135 observations deleted due to missingness)
## Multiple R-squared:  0.007964,   Adjusted R-squared:  0.007838
## F-statistic: 62.94 on 1 and 7840 DF,  p-value: 2.426e-15
```

#Renaming class_id to easier names to read on graph

```
checkpoints_pulse <- checkpoints_pulse %>%  mutate(class_id = case_when(class_id == "0089dedf-6316-4c32-
                              class_id == "074123e7-cd90-4500-86fe-286aaa733bf5" ~ "C-02",
                              class_id == "0d546479-6f77-4477-9c7e-365cd36c97eb" ~ "C-03",
                              class_id == "1020418a-3eeb-4251-88f7-150c8fe00a56" ~ "C-04",
                              class_id == "103f5ce8-9e95-4916-815e-9f821d274a59" ~ "C-05",
                              class_id == "1cca9f91-5c4a-4e1a-8e0e-293b070dfd6f" ~ "C-06",
                              class_id == "20bd524c-bb2d-4b74-a419-929475b91d94" ~ "C-07",
                              class_id == "2294d558-6f5d-41c5-8d28-7b5280970f95" ~ "C-08",
                              class_id == "3631cec9-51d3-4237-906f-a142a715be51" ~ "C-09",
                              class_id == "40e49bfa-f6cb-42fa-a3a4-b23592b799ec" ~ "C-10",
                              class_id == "4a3b5b2c-ef0f-4121-96f4-fd8a42764836" ~ "C-11",
                              class_id == "51711479-441b-4c02-aef7-517aca63a53f" ~ "C-12",
                              class_id == "52619962-72f6-4716-9c64-1c06fe10f739" ~ "C-13",
                              class_id == "552ede8f-6b54-426d-8d29-abdc43a668cb" ~ "C-14",
                              class_id == "5bd961c4-659c-40a7-a685-6735189f2b65" ~ "C-15",
                              class_id == "60e05fa5-c986-4973-9833-16238720b727" ~ "C-16",
                              class_id == "65246c1e-a176-4760-acb5-a320a9b7b2fe" ~ "C-17",
                              class_id == "686478e7-82ac-4e6c-a3ec-2da0076ef868" ~ "C-18",
                              class_id == "79662249-02f6-48d8-aa99-1e1c0aeea77d" ~ "C-19",
                              class_id == "7a987176-7e55-45b5-a715-7f56c59d5f49" ~ "C-20",
                              class_id == "822d72d9-0c18-47a0-99fc-7223b4fd22f5" ~ "C-21",
                              class_id == "8589cd83-192c-44c8-b649-cd848e519530" ~ "C-22",
                              class_id == "94da41a4-f9f8-4225-bf41-42db737850b9" ~ "C-23",
                              class_id == "97c61e74-5a20-4cf5-bf67-8f8db750d0e7" ~ "C-24",
                              class_id == "98119d92-8cc6-416a-972c-630351726223" ~ "C-25",
                              class_id == "9bdf8bfc-9998-4fd8-85d2-70c91cf94891" ~ "C-26",
                              class_id == "9fad0c9e-9d3d-4eed-ada6-3959bd6d712c" ~ "C-27",
                              class_id == "afcb6b4e-a0c0-46ce-b38c-c96329c91471" ~ "C-28",
                              class_id == "b1421b49-4026-4c61-9786-d4ef110c8db3" ~ "C-29",
                              class_id == "b16b895d-ca1d-4330-a36d-c43fb33436e5" ~ "C-30",
                              class_id == "bc650f4f-11f0-439a-a90a-47726724c811" ~ "C-31",
                              class_id == "bcae937d-c95f-436c-ac0f-d4a5e995de19" ~ "C-32",
                              class_id == "c09145c1-d635-41ae-b881-17ab46895fe4" ~ "C-33",
                              class_id == "c1168ee3-7ac8-4fdc-af0e-e375ad0629fe" ~ "C-34",
                              class_id == "c7008a64-b43c-4eb4-bebf-07b08b9894ad" ~ "C-35",
                              class_id == "cc1ffb2e-5555-4109-8ad8-2d49cb54ad10" ~ "C-36",
                              class_id == "d0b4f5e2-6d8f-4828-91cd-3f4714b821b0" ~ "C-37",
                              class_id == "fe8c4185-7e8d-48eb-bf0e-85562e060d5d" ~ "C-38"))

ggplot(checkpoints_pulse, aes(y = response, fill = construct)) +
  theme_minimal() +
  geom_boxplot(outliers = F) +
  facet_wrap(~class_id)  +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1", n = 4, type = "discrete"))
```

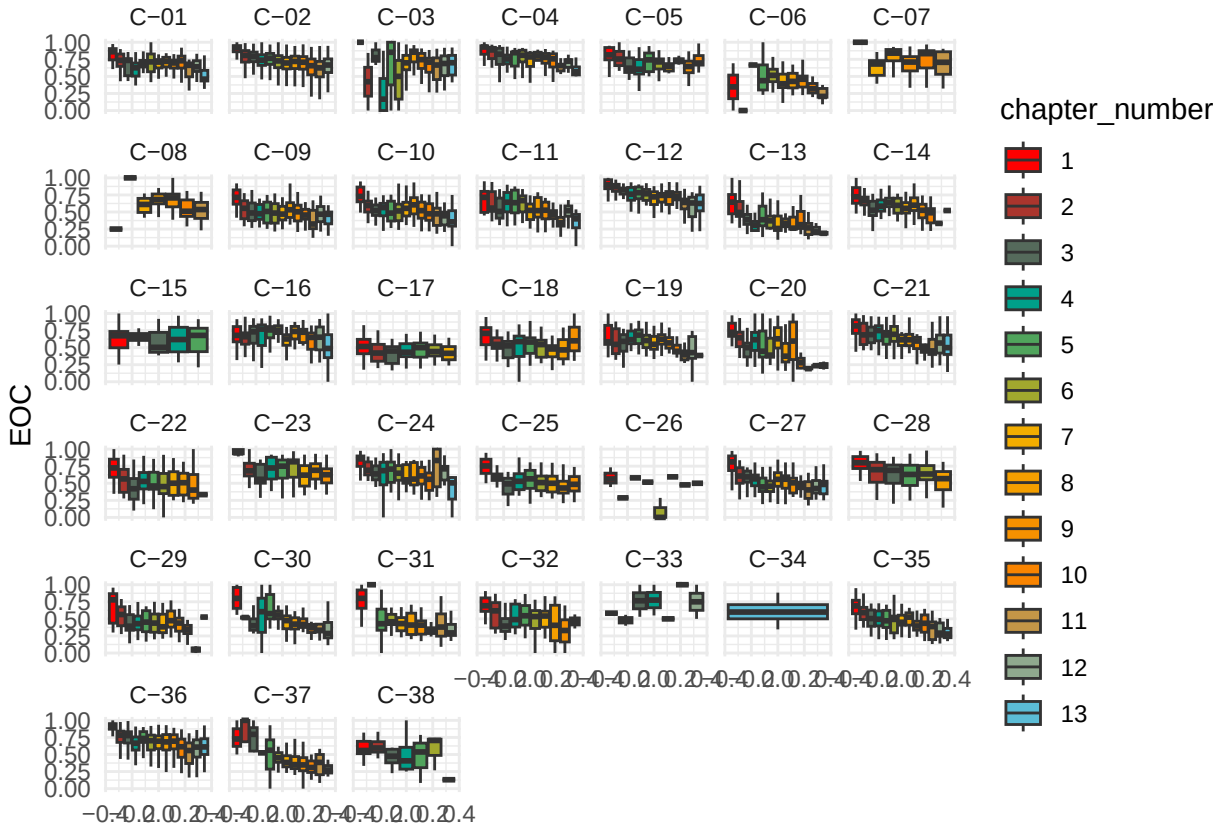
```
## Warning: Removed 32386 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



code for EOC important stuff

```
eoc <- checkpoints_eoc %>% filter(book == "College / Statistics and Data Science (ABC)")
eoc <- eoc %>% filter(EOC != "")
eoc <- eoc %>% mutate(avg_try = n_attempt/n_possible)

eoc <- eoc %>%  mutate(class_id = case_when(class_id == "0089dedf-6316-4c32-a38c-d48dfafed882" ~ "C-01"
                                class_id == "074123e7-cd90-4500-86fe-286aaa733bf5" ~ "C-02",
                                class_id == "0d546479-6f77-4477-9c7e-365cd36c97eb" ~ "C-03",
                                class_id == "1020418a-3eeb-4251-88f7-150c8fe00a56" ~ "C-04",
                                class_id == "103f5ce8-9e95-4916-815e-9f821d274a59" ~ "C-05",
                                class_id == "1cca9f91-5c4a-4e1a-8e0e-293b070dfd6f" ~ "C-06",
                                class_id == "20bd524c-bb2d-4b74-a419-929475b91d94" ~ "C-07",
                                class_id == "2294d558-6f5d-41c5-8d28-7b5280970f95" ~ "C-08",
                                class_id == "3631cec9-51d3-4237-906f-a142a715be51" ~ "C-09",
                                class_id == "40e49bfa-f6cb-42fa-a3a4-b23592b799ec" ~ "C-10",
                                class_id == "4a3b5b2c-ef0f-4121-96f4-fd8a42764836" ~ "C-11",
                                class_id == "51711479-441b-4c02-aef7-517aca63a53f" ~ "C-12",
                                class_id == "52619962-72f6-4716-9c64-1c06fe10f739" ~ "C-13",
                                class_id == "552ede8f-6b54-426d-8d29-abdc43a668cb" ~ "C-14",
                                class_id == "5bd961c4-659c-40a7-a685-6735189f2b65" ~ "C-15",
                                class_id == "60e05fa5-c986-4973-9833-16238720b727" ~ "C-16",
                                class_id == "65246c1e-a176-4760-acb5-a320a9b7b2fe" ~ "C-17",
                                class_id == "686478e7-82ac-4e6c-a3ec-2da0076ef868" ~ "C-18",
                                class_id == "79662249-02f6-48d8-aa99-1e1c0aeea77d" ~ "C-19",
                                class_id == "7a987176-7e55-45b5-a715-7f56c59d5f49" ~ "C-20",
```

```
                              class_id == "822d72d9-0c18-47a0-99fc-7223b4fd22f5" ~ "C-21",
                              class_id == "8589cd83-192c-44c8-b649-cd848e519530" ~ "C-22",
                              class_id == "94da41a4-f9f8-4225-bf41-42db737850b9" ~ "C-23",
                              class_id == "97c61e74-5a20-4cf5-bf67-8f8db750d0e7" ~ "C-24",
                              class_id == "98119d92-8cc6-416a-972c-630351726223" ~ "C-25",
                              class_id == "9bdf8bfc-9998-4fd8-85d2-70c91cf94891" ~ "C-26",
                              class_id == "9fad0c9e-9d3d-4eed-ada6-3959bd6d712c" ~ "C-27",
                              class_id == "afcb6b4e-a0c0-46ce-b38c-c96329c91471" ~ "C-28",
                              class_id == "b1421b49-4026-4c61-9786-d4ef110c8db3" ~ "C-29",
                              class_id == "b16b895d-ca1d-4330-a36d-c43fb33436e5" ~ "C-30",
                              class_id == "bc650f4f-11f0-439a-a90a-47726724c811" ~ "C-31",
                              class_id == "bcae937d-c95f-436c-ac0f-d4a5e995de19" ~ "C-32",
                              class_id == "c09145c1-d635-41ae-b881-17ab46895fe4" ~ "C-33",
                              class_id == "c1168ee3-7ac8-4fdc-af0e-e375ad0629fe" ~ "C-34",
                              class_id == "c7008a64-b43c-4eb4-bebf-07b08b9894ad" ~ "C-35",
                              class_id == "cc1ffb2e-5555-4109-8ad8-2d49cb54ad10" ~ "C-36",
                              class_id == "d0b4f5e2-6d8f-4828-91cd-3f4714b821b0" ~ "C-37",
                              class_id == "fe8c4185-7e8d-48eb-bf0e-85562e060d5d" ~ "C-38"))

eoc_temp <- eoc
eoc_temp$chapter_number <-  as.factor(eoc_temp$chapter_number)

#EOC with fill by chapter_number
ggplot(eoc_temp, aes(y = EOC, fill = chapter_number)) +
  theme_minimal() +
  geom_boxplot(outliers = F) +
  facet_wrap(~class_id)  +
  scale_fill_manual(values = wes_palette(name = "Darjeeling1", n = 13, type = "continuous"))
```



#wordcloud

```
library(tidyverse)

pass_responses <- responses %>%
  mutate(points_possible = as.numeric(points_possible)) %>%
  mutate(points_earned = as.numeric(points_earned)) %>%
  filter(!is.na(points_earned)) %>%
  filter(!is.na(points_possible)) %>%
  mutate(perc_score = points_possible/points_earned) %>%
  filter(perc_score > .6) %>%
  filter(institution_id == "97aebe75-a051-4bff-a2c0-1d53eb5d9498")
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `points_possible = as.numeric(points_possible)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `points_earned = as.numeric(points_earned)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
#install.packages("wordcloud")
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
#install.packages("RColorBrewer")
library(RColorBrewer)
```

```r
#install.packages("wordcloud2")
library(wordcloud2)

#install.packages("tm")
library(tm)
```

```
## Warning: package 'tm' was built under R version 4.2.3
```

```
## Loading required package: NLP
```

```
##
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```r
#fc5f1b1b-2aeb-4e09-93fc-06fdac0d8030
# Making DF for word clouds

# Pre word cloud
corpus = Corpus(VectorSource(pass_responses$response))

corpus <- corpus %>%
  tm_map(removeNumbers) %>%
  tm_map(removePunctuation) %>%
  tm_map(stripWhitespace) %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removeWords, stopwords("english")) %>%
  tm_map(removeWords, stopwords("SMART"))
```

```
## Warning in tm_map.SimpleCorpus(., removeNumbers): transformation drops
## documents
```

```
## Warning in tm_map.SimpleCorpus(., removePunctuation): transformation drops
## documents
```

```
## Warning in tm_map.SimpleCorpus(., stripWhitespace): transformation drops
## documents
```

```
## Warning in tm_map.SimpleCorpus(., content_transformer(tolower)): transformation
## drops documents
```

```
## Warning in tm_map.SimpleCorpus(., removeWords, stopwords("english")):
## transformation drops documents
```

```
## Warning in tm_map.SimpleCorpus(., removeWords, stopwords("SMART")):
## transformation drops documents
```

```r
tdm = TermDocumentMatrix(corpus) %>%
  as.matrix()

words = sort(rowSums(tdm), decreasing = TRUE)

pre_WCdf = data.frame(words = names(words), freq = words)


# Color Palettes
```

```
pre_WCcolors = c("#510C76", "#00A8E2", "#87714D")
pre_WCbkgd = "#FFFFFF"
post_WCcolors = c("#FFFFFF", "#510C76", "#87714D")
post_WCbkgd = "#00A8E2"

#rm unneeded vars
rm(corpus, tdm, words)

WC_Pre <- wordcloud2(pre_WCdf,
           color = rep_len(pre_WCcolors, nrow(pre_WCdf)),
           backgroundColor = pre_WCbkgd,
           fontFamily = "AppleMyungjo",
           size = .62,
           rotateRatio = 0)

#Final wordcloud
WC_Pre
```