

eda

Yining Chen

2023-12-03

```
aids <- read.csv('data/AIDS_Clinical_Trials_Group175.csv')
aids <- aids%>%
  mutate(cid = as.factor(cid),

          trt = as.factor(trt), treat=as.factor(treat),

          hemo = as.factor(hemo),
          homo = as.factor(homo),
          drugs = as.factor(drugs),
          race = as.factor(race),
          gender = as.factor(gender),
          str2 = as.factor(str2),
          symptom = as.factor(symptom))
```

```
aids <- aids%>%
  mutate(cid=factor(cid,
    levels = c(0,1),
    labels = c("Censoring",
               "Failure")),
  treat=factor(treat,
    levels = c(0,1),
    labels = c("ZDV only",
               "Others")),
  hemo=factor(hemo,
    levels = c(0,1),
    labels = c("No",
               "Yes")),
  homo=factor(homo,
    levels = c(0,1),
    labels = c("No",
               "Yes")),
  gender=factor(gender,
    levels = c(0,1),
    labels = c("Female",
               "Male")),
  race=factor(race,
    levels = c(0,1),
    labels = c("White",
               "Non-white")),
  drugs=factor(drugs,
    levels = c(0,1),
    labels = c("No",
               "Yes")),
```

```

symptom=factor(symptom,
levels = c(0,1),
labels = c("No",
           "Yes"))
)

```

sex age race hemo homo drug kanor symptom cd40

```

explanatory = c("age", "hemo", "homo", "race", "gender", "drugs", "karnof", "cd40", "symptom", "cd80", "wtkg")
dependent = "treat"
baseline <- aids %>%
  mutate(
    cd80 = ff_label(cd80, "CD8 Count"),
    wtkg = ff_label(wtkg, "Weight"),
    gender = ff_label(gender, "Gender"),
    hemo = ff_label(hemo, "Hemophilia"),
    homo = ff_label(homo, "Homosexuality"),
    race = ff_label(race, "Race"),
    drugs = ff_label(drugs, "History of IV drug use "),
    karnof = ff_label(karnof, "Karnofsky score of 100"),
    cd40 = ff_label(cd40, "CD4 count"),
    age = ff_label(age, "Age"),
    symptom = ff_label(symptom, "Symptomatic infection"),
    treat = ff_label(treat, "Treatment")
  )%>%
  summary_factorlist(dependent, explanatory, column = TRUE, total_col = TRUE, col_totals_prefix = "N=", a
baseline

```

##	Dependent: Treatment		ZDV only	Others	Total
##	Age	Mean (SD)	35.2 (8.9)	35.3 (8.7)	35.2 (8.7)
##	Hemophilia	No	490 (92.1)	1469 (91.4)	1959 (91.6)
##		Yes	42 (7.9)	138 (8.6)	180 (8.4)
##	Homosexuality	No	191 (35.9)	534 (33.2)	725 (33.9)
##		Yes	341 (64.1)	1073 (66.8)	1414 (66.1)
##	Race	White	376 (70.7)	1146 (71.3)	1522 (71.2)
##		Non-white	156 (29.3)	461 (28.7)	617 (28.8)
##	Gender	Female	100 (18.8)	268 (16.7)	368 (17.2)
##		Male	432 (81.2)	1339 (83.3)	1771 (82.8)
##	History of IV drug use	No	469 (88.2)	1389 (86.4)	1858 (86.9)
##		Yes	63 (11.8)	218 (13.6)	281 (13.1)
##	Karnofsky score of 100	70	4 (0.8)	5 (0.3)	9 (0.4)
##		80	17 (3.2)	63 (3.9)	80 (3.7)
##		90	197 (37.0)	590 (36.7)	787 (36.8)
##		100	314 (59.0)	949 (59.1)	1263 (59.0)
##	CD4 count	Mean (SD)	353.2 (114.1)	349.6 (120.0)	350.5 (118.6)
##	Symptomatic infection	No	443 (83.3)	1326 (82.5)	1769 (82.7)
##		Yes	89 (16.7)	281 (17.5)	370 (17.3)
##	CD8 Count	Mean (SD)	987.2 (475.2)	986.4 (482.0)	986.6 (480.2)
##	Weight	Mean (SD)	76.1 (13.2)	74.8 (13.3)	75.1 (13.3)

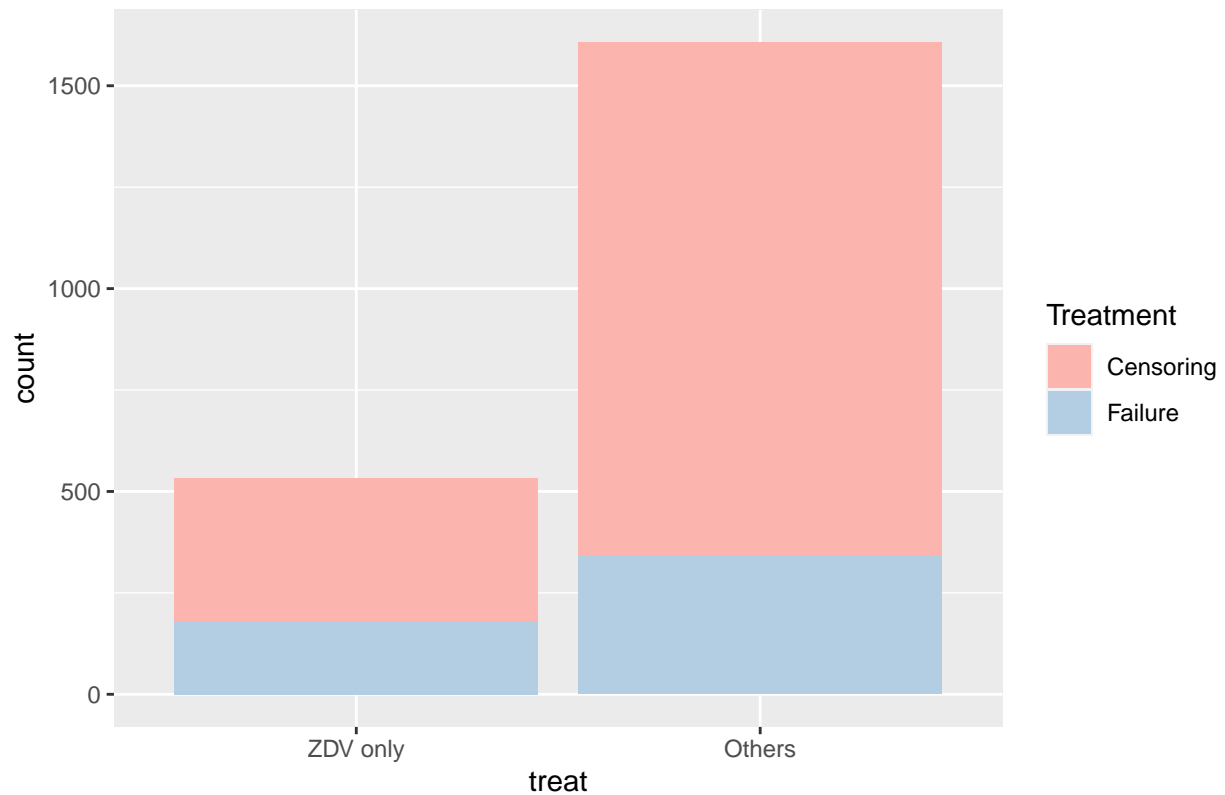
```
kable(baseline,caption = "Base-Line Characteristics of the Patients According to the Treatment Indicator")
```

Table 1: Base-Line Characteristics of the Patients According to the Treatment Indicator

Dependent: Treatment		ZDV only	Others	Total
Age	Mean (SD)	35.2 (8.9)	35.3 (8.7)	35.2 (8.7)
Hemophilia	No	490 (92.1)	1469 (91.4)	1959 (91.6)
	Yes	42 (7.9)	138 (8.6)	180 (8.4)
Homosexuality	No	191 (35.9)	534 (33.2)	725 (33.9)
	Yes	341 (64.1)	1073 (66.8)	1414 (66.1)
Race	White	376 (70.7)	1146 (71.3)	1522 (71.2)
	Non-white	156 (29.3)	461 (28.7)	617 (28.8)
Gender	Female	100 (18.8)	268 (16.7)	368 (17.2)
	Male	432 (81.2)	1339 (83.3)	1771 (82.8)
History of IV drug use	No	469 (88.2)	1389 (86.4)	1858 (86.9)
	Yes	63 (11.8)	218 (13.6)	281 (13.1)
Karnofsky score of 100	70	4 (0.8)	5 (0.3)	9 (0.4)
	80	17 (3.2)	63 (3.9)	80 (3.7)
	90	197 (37.0)	590 (36.7)	787 (36.8)
	100	314 (59.0)	949 (59.1)	1263 (59.0)
CD4 count	Mean (SD)	353.2 (114.1)	349.6 (120.0)	350.5 (118.6)
Symptomatic infection	No	443 (83.3)	1326 (82.5)	1769 (82.7)
	Yes	89 (16.7)	281 (17.5)	370 (17.3)
CD8 Count	Mean (SD)	987.2 (475.2)	986.4 (482.0)	986.6 (480.2)
Weight	Mean (SD)	76.1 (13.2)	74.8 (13.3)	75.1 (13.3)

```
ggplot(data=aids, aes(x=treat, fill=cid)) +
  geom_bar() +
  scale_fill_brewer(palette="Pastel1") +
  ggtitle("Figure 1: Distribution of Treatment VS. Patient Status") +
  guides(fill = guide_legend(title = "Treatment"))
```

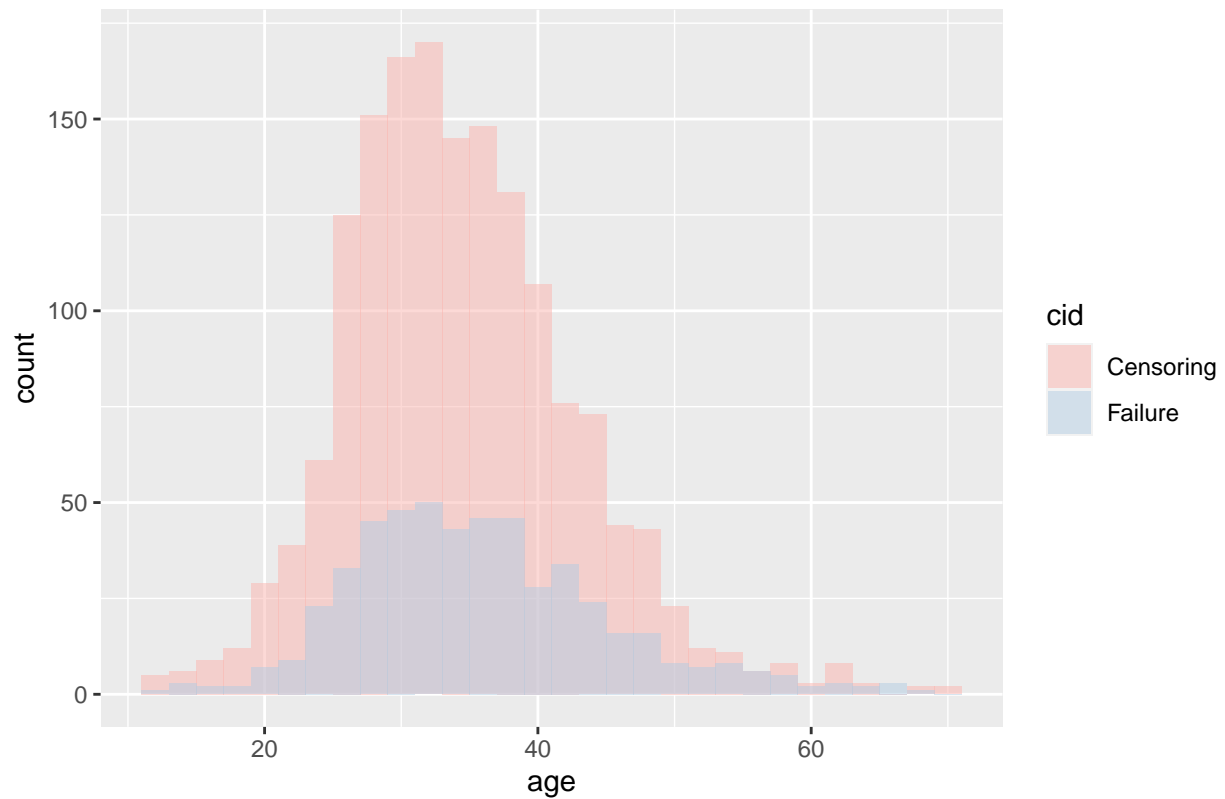
Figure 1: Distribution of Treatment VS. Patient Status



```
ggplot(aids, aes(x=age, fill=cid)) +  
  geom_histogram(alpha=0.5, position="identity")+  
  scale_fill_brewer(palette = "Pastell")+  
  ggtitle("Figure 2: Distribution of Patient Age VS. Status")
```

## 'stat\_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

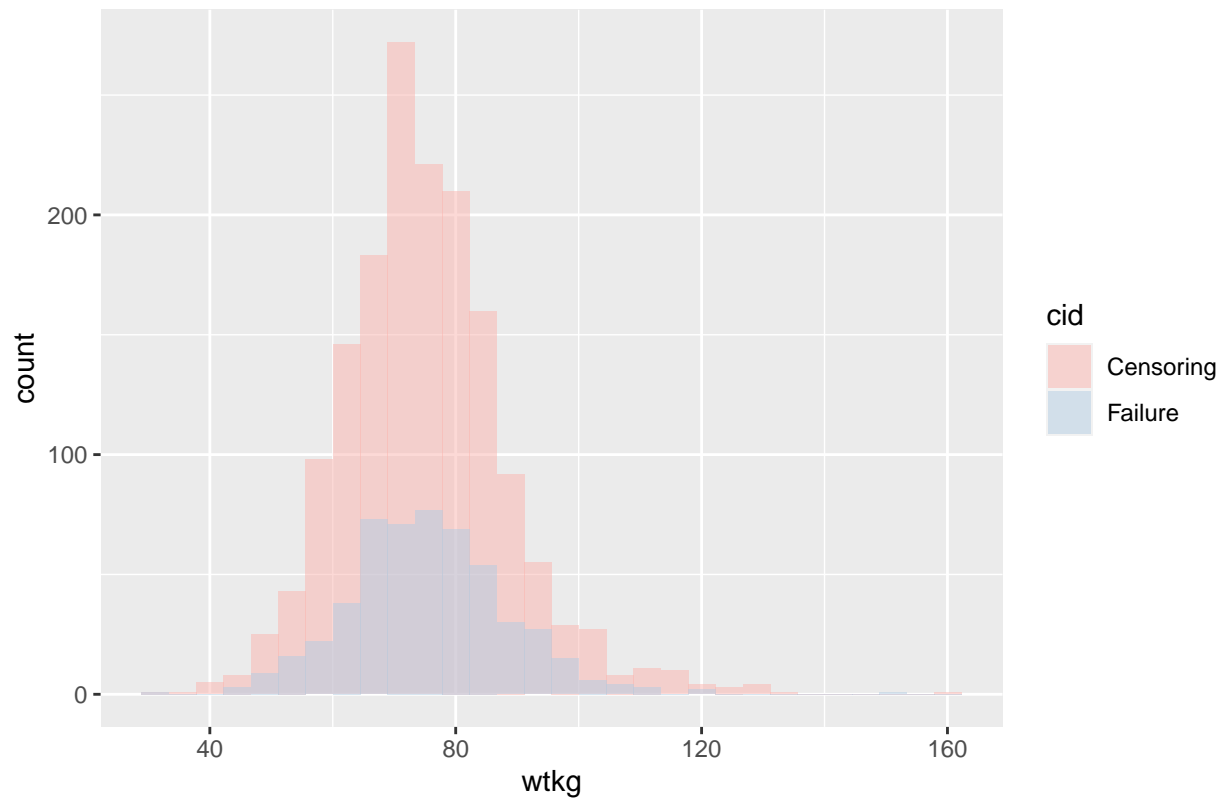
Figure 2: Distribution of Patient Age VS. Status



```
ggplot(aids, aes(x=wtkg, fill=cid)) +  
  geom_histogram(alpha=0.5, position="identity")+  
  scale_fill_brewer(palette = "Pastell")+  
  ggtitle("Figure 2: Distribution of Patient Weight VS. Status")
```

## 'stat\_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

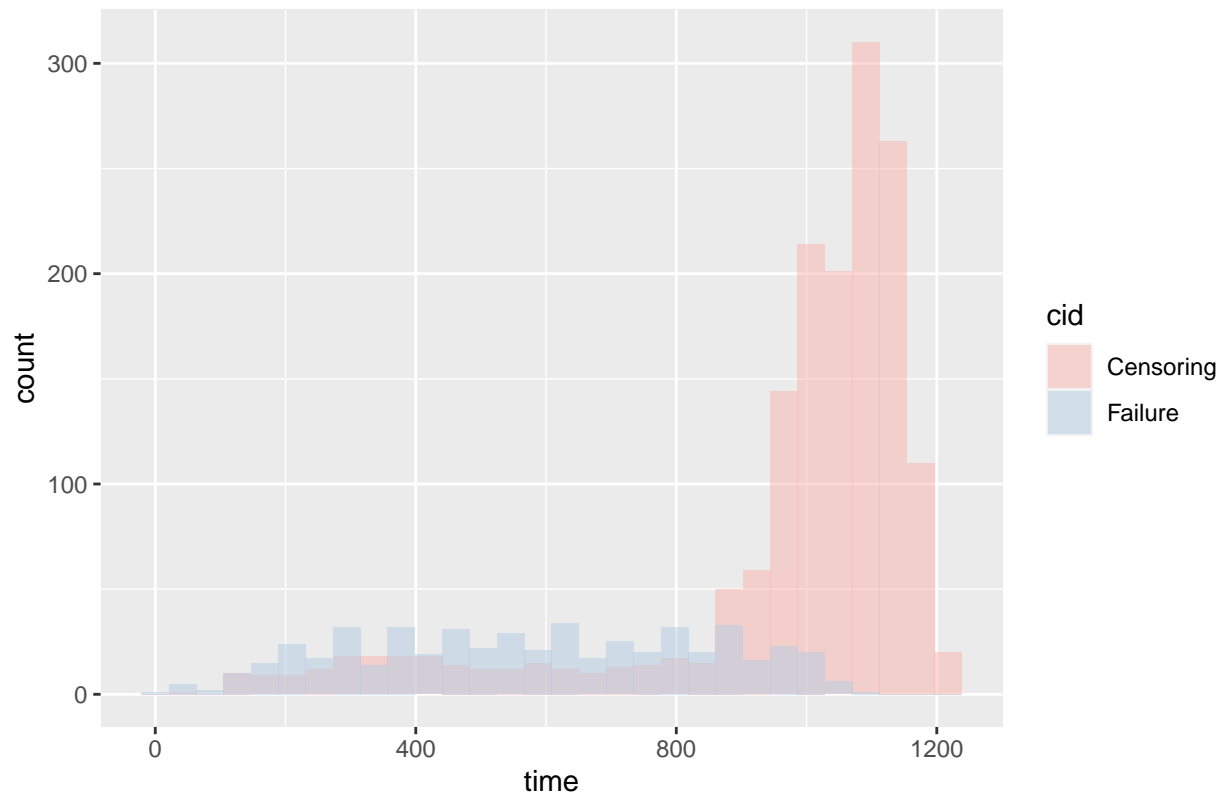
Figure 2: Distribution of Patient Weight VS. Status



```
ggplot(aids, aes(x=time, fill=cid)) +  
  geom_histogram(alpha=0.5, position="identity")+  
  scale_fill_brewer(palette = "Pastell")+  
  ggtitle("Figure 3: Distribution of Time to Failure or Censoring")
```

## 'stat\_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

Figure 3: Distribution of Time to Failure or Censoring



```
p1 <- ggplot(aids,aes(x= treat, y=cd40,color=treat)) + geom_boxplot(show.legend = FALSE)+labs(x="")+scale_y_continuous(limits = c(0, 1000))
theme_minimal()

p2 <- ggplot(aids,aes(x= treat, y=cd420,color=treat)) + geom_boxplot(show.legend = FALSE)+labs(x="")+scale_y_continuous(limits = c(0, 1000))
theme_minimal()

p <- ggpubr::ggarrange(p1, p2, ncol=2,nrow = 1,common.legend = TRUE)
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_boxplot()').
```

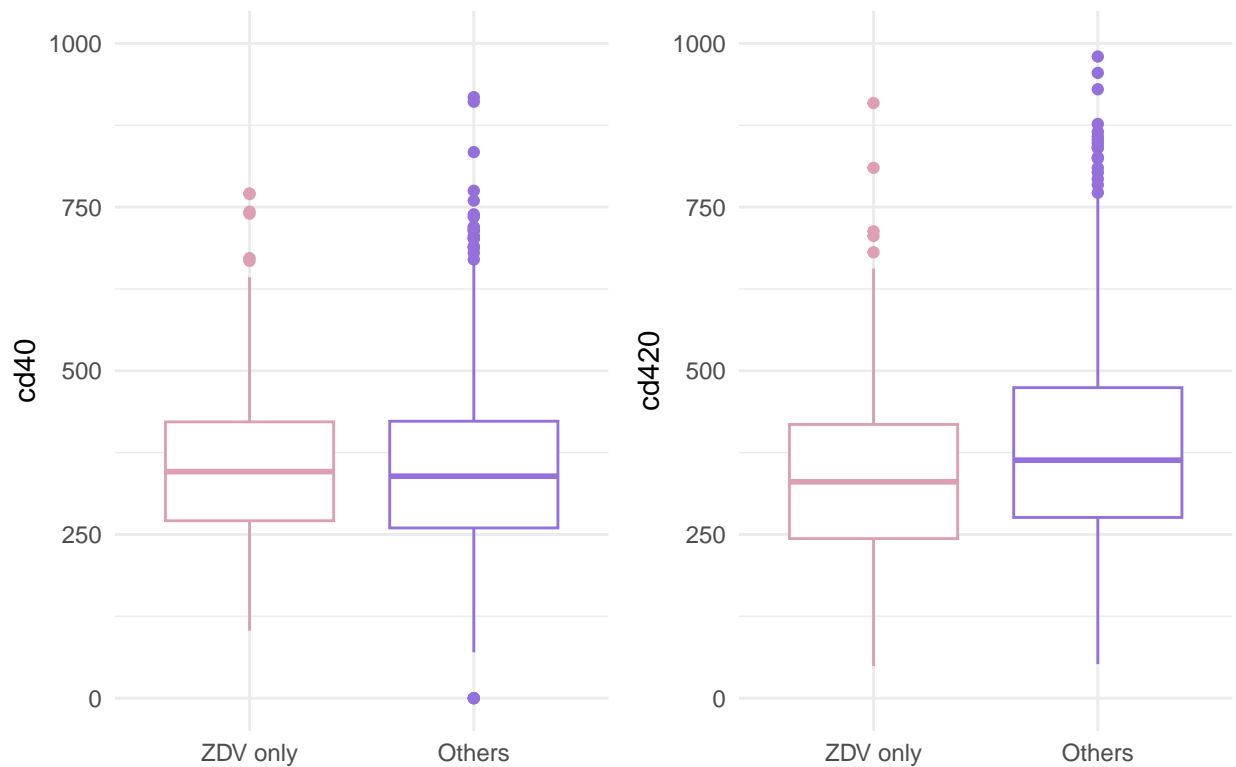
```
## Warning: Removed 3 rows containing non-finite values ('stat_boxplot()').
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_boxplot()').
```

```
## Warning: Removed 3 rows containing non-finite values ('stat_boxplot()').
```

```
ggpubr::annotate_figure(p, top = ggpubr::text_grob("Figure 4: CD4 Count Change VS. Treatment Groups", c
```

Figure 4: CD4 Count Change VS. Treatment Groups



```
p3 <- ggplot(aids,aes(x= treat, y=cd80,color=treat)) + geom_boxplot(show.legend = FALSE)+labs(x="")+scale_y_continuous(limits = c(0, 4000))
theme_minimal()

p4 <- ggplot(aids,aes(x= treat, y=cd820,color=treat)) + geom_boxplot(show.legend = FALSE)+labs(x="")+scale_y_continuous(limits = c(0, 4000))
theme_minimal()

p<- ggarrange(p3,p4,ncol=2,align = "hv",nrow = 1,common.legend = TRUE)
```

```
## Warning: Removed 2 rows containing non-finite values ('stat_boxplot()').
## Removed 2 rows containing non-finite values ('stat_boxplot()').
## Removed 2 rows containing non-finite values ('stat_boxplot()').
## Removed 2 rows containing non-finite values ('stat_boxplot()').
```

```
annotate_figure(p, top = text_grob("Figure 5: CD8 Count Change VS. Treatment Groups", color = "#0F2540"
```



**Figure 5: CD8 Count Change VS. Treatment Groups**

