Number of samples in TCGA lung cancer cohorts

load R packages

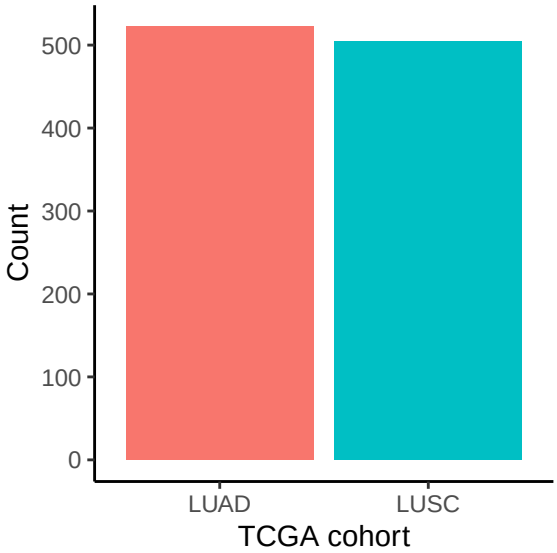
```
library(tidyverse)
```

load samples metadata

```
luad_metadata <- read_tsv("./data/processed/metadata/tcga_luad_metadata_pancancer_atlas.txt")
lusc_metadata <- read_tsv("./data/processed/metadata/tcga_lusc_metadata_pancancer_atlas.txt")

tcga_lung <- bind_rows(luad_metadata, lusc_metadata)
```

number of primary tumours in both cohorts

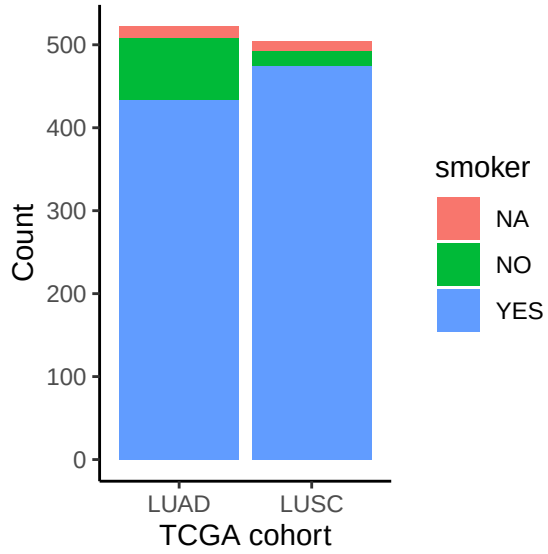
```
samples_by_cohort <- tcga_lung %>%
  ggplot(mapping = aes(x = tcga_cohort, fill = tcga_cohort)) +
  geom_bar(show.legend = F) +
  theme_classic() +
  labs(x = "TCGA cohort", y = "Count")

samples_by_cohort
```



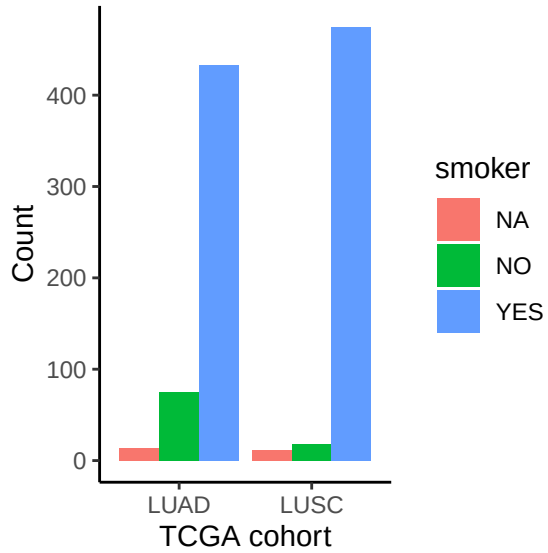
number of smokers and nonsmokers in both cohorts

```
smokers_by_cohort <- tcga_lung %>%
  mutate(smoker = str_replace_na(smoker)) %>%
  ggplot(mapping = aes(x = tcga_cohort, fill = smoker)) +
  geom_bar() +
  theme_classic() +
```

```
  labs(x = "TCGA cohort", y = "Count")

smokers_by_cohort
```

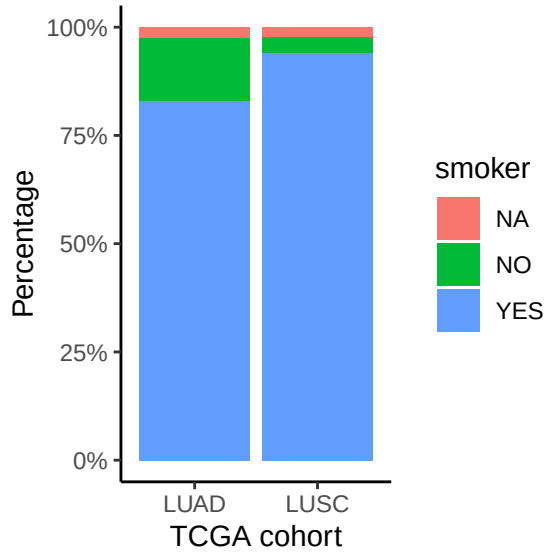



```
smokers_by_cohort <- tcga_lung %>%
  mutate(smoker = str_replace_na(smoker)) %>%
  ggplot(mapping = aes(x = tcga_cohort, fill = smoker)) +
  geom_bar(position = "dodge") +
  theme_classic() +
  labs(x = "TCGA cohort", y = "Count")

smokers_by_cohort
```



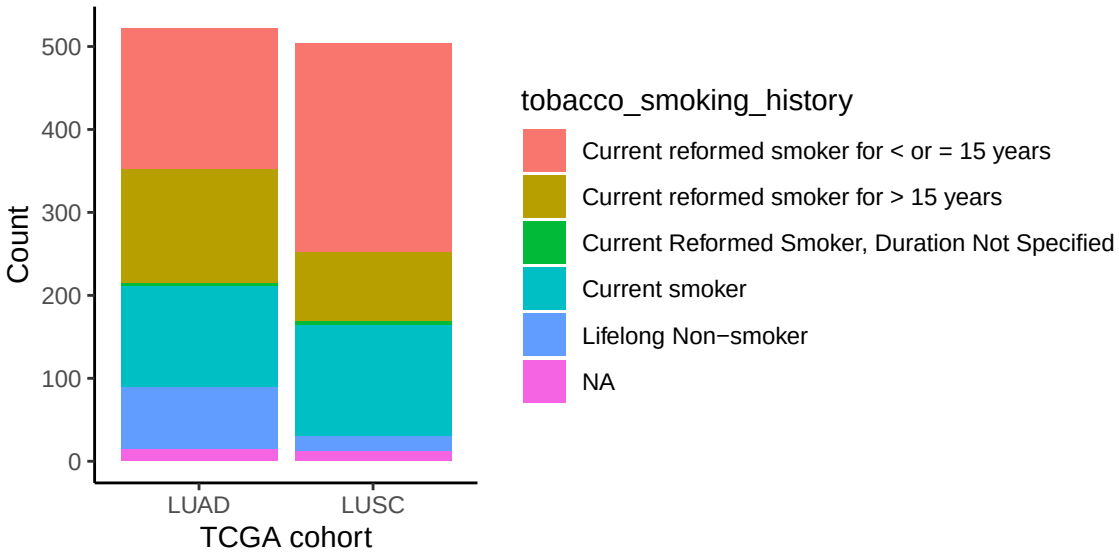

```
smokers_by_cohort <- tcga_lung %>%
  mutate(smoker = str_replace_na(smoker)) %>%
  ggplot(mapping = aes(x = tcga_cohort, fill = smoker)) +
```

```
  geom_bar(position = "fill") +
  theme_classic() +
  scale_y_continuous(labels = ~ paste0(.x*100, "%")) +
  labs(x = "TCGA cohort", y = "Percentage")

smokers_by_cohort
```

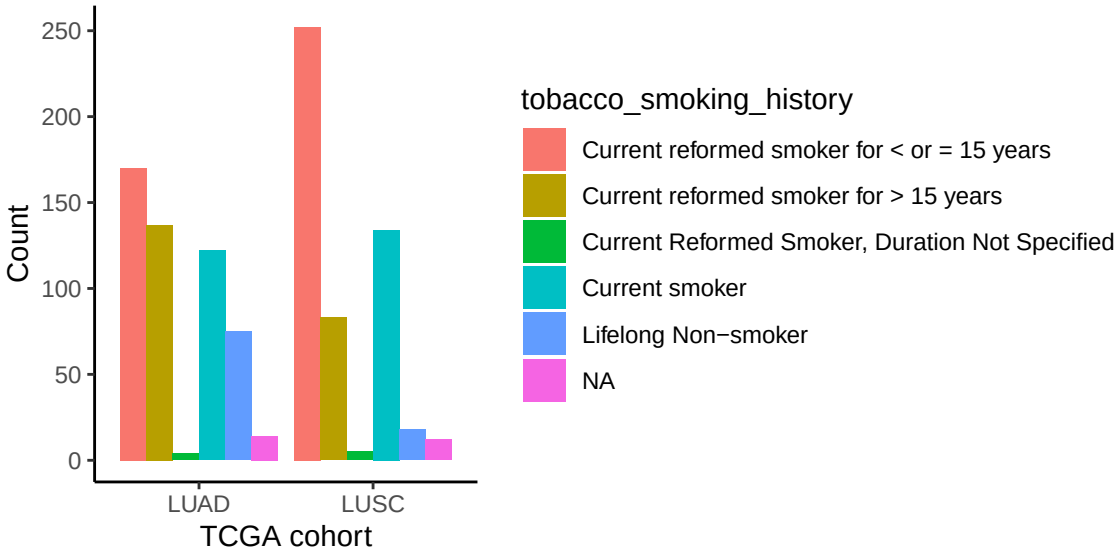


number of smokers stratified by smoking history

```
smokers_by_cohort <- tcga_lung %>%
  mutate(tobacco_smoking_history = str_replace_na(tobacco_smoking_history)) %>%
  ggplot(mapping = aes(x = tcga_cohort, fill = tobacco_smoking_history)) +
  geom_bar() +
  theme_classic() +
  labs(x = "TCGA cohort", y = "Count")

smokers_by_cohort
```




```
smokers_by_cohort <- tcga_lung %>%
  mutate(tobacco_smoking_history = str_replace_na(tobacco_smoking_history)) %>%
  ggplot(mapping = aes(x = tcga_cohort, fill = tobacco_smoking_history)) +
  geom_bar(position = "dodge") +
  theme_classic() +
  labs(x = "TCGA cohort", y = "Count")

smokers_by_cohort
```




```
smokers_by_cohort <- tcga_lung %>%
  mutate(tobacco_smoking_history = str_replace_na(tobacco_smoking_history)) %>%
  ggplot(mapping = aes(x = tcga_cohort, fill = tobacco_smoking_history)) +
  geom_bar(position = "fill") +
  theme_classic() +
  scale_y_continuous(labels = ~ paste0(.x*100, "%")) +
  labs(x = "TCGA cohort", y = "Percentage")
```

```
smokers_by_cohort
```