Secondary structure of TCR CDR regions

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(stringr)
```

Load TCR V-D-J mapping for our PDBs

```
df.meta = read.table("tcr.annotations.txt", header = T, sep = "\t") %>%
  filter(region %in% c("CDR1", "CDR2", "CDR3"))

df.meta$tcr_gene = substr(df.meta$v_allele, 0, 3)
```

Load data processed with STRIDE software, see https://en.wikipedia.org/wiki/STRIDE and https://en.wikipedia.org/wiki/DSSP_(hydrogen_bond_estimation_algorithm). See DSSP classification section in https://en.wikipedia.org/wiki/Protein_secondary_structure for glossary

```
df.ss = data.frame()

for (f in list.files('data/')) {
  .cmd = paste0('zcat data/',  f, ' | grep "^ASG"')

  .df.ss.tmp = read.table(pipe(.cmd), header = F) %>%
               select(V2, V3, V5, V7, V11)

  colnames(.df.ss.tmp) = c("res", "pdb_chain_id", "pos", "ss", "pdb_id")

  .df.ss.tmp$pdb_id = tolower(.df.ss.tmp$pdb_id)
  df.ss = rbind(df.ss, .df.ss.tmp)
}

df.ss$pos = as.integer(df.ss$pos)
df.ss$ss = as.factor(df.ss$ss)
```

Merge with CDR1-3 mapping

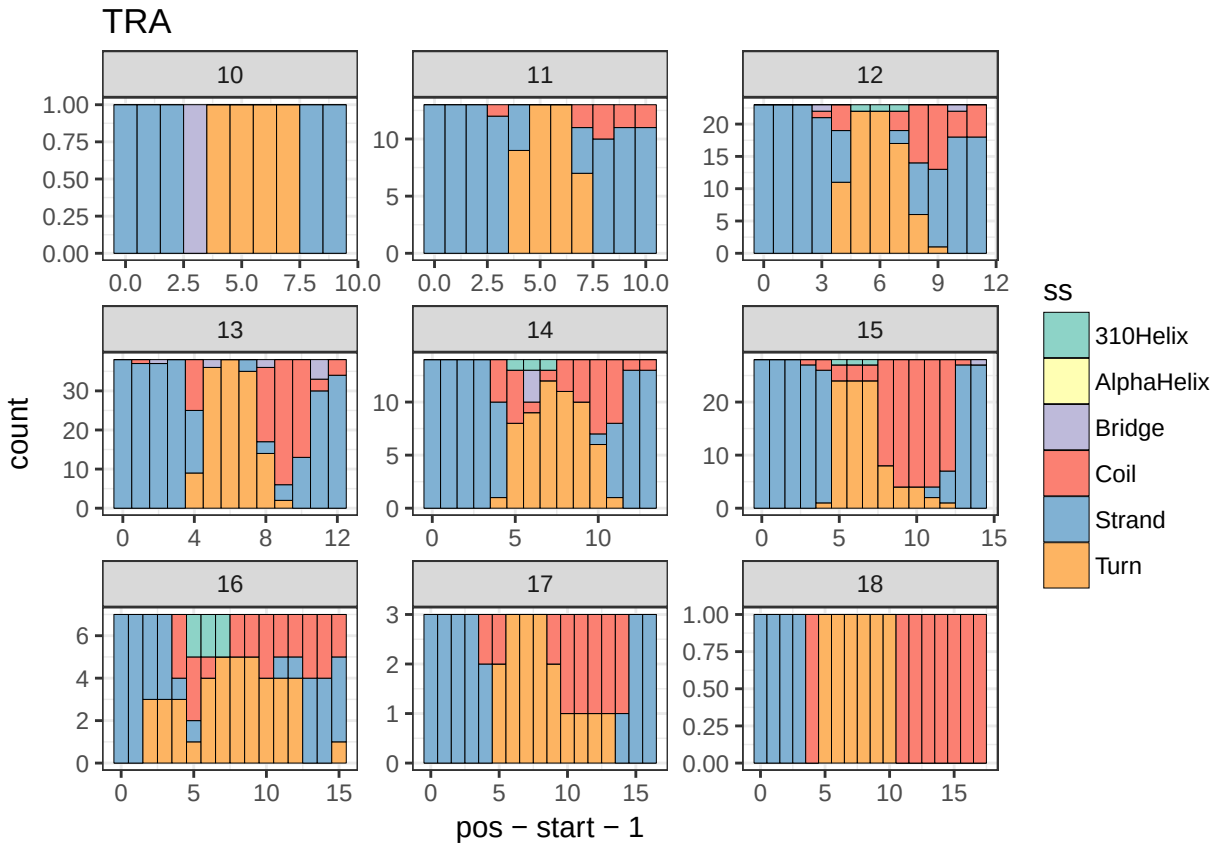
```
df.ss.1 = merge(df.meta, df.ss, by = c("pdb_id", "pdb_chain_id"), all.y = T) %>%
 filter(pos >= start+1 & pos <= end) %>% droplevels()

df.ss.1$len_tcr = with(df.ss.1, end - start)
```

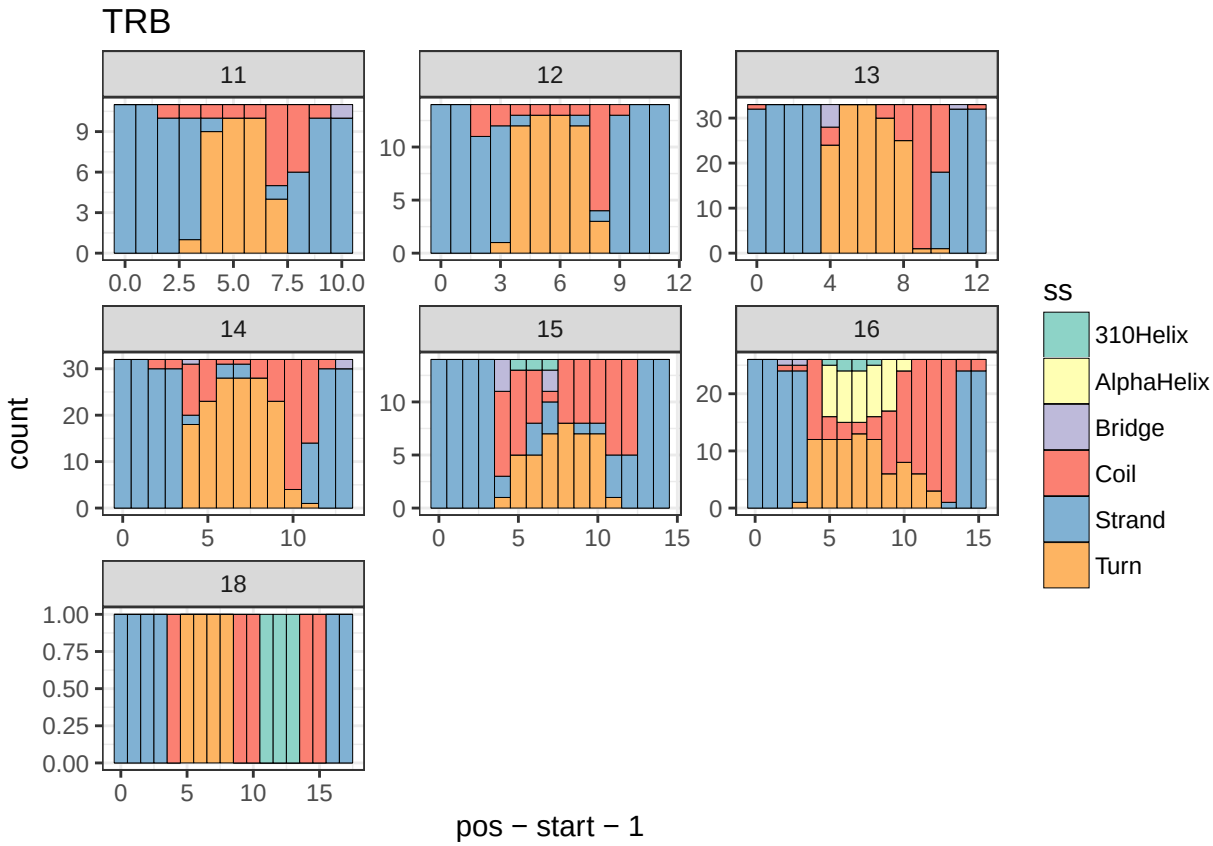
Distribution of various secondary structure classifications for CDR3

```
ggplot(df.ss.1 %>% filter(region=="CDR3" & len_tcr > 5 & tcr_gene=="TRA"), aes(x=pos - start - 1, fill =
  geom_histogram(binwidth = 1, color = "black", size = 0.1) +
```

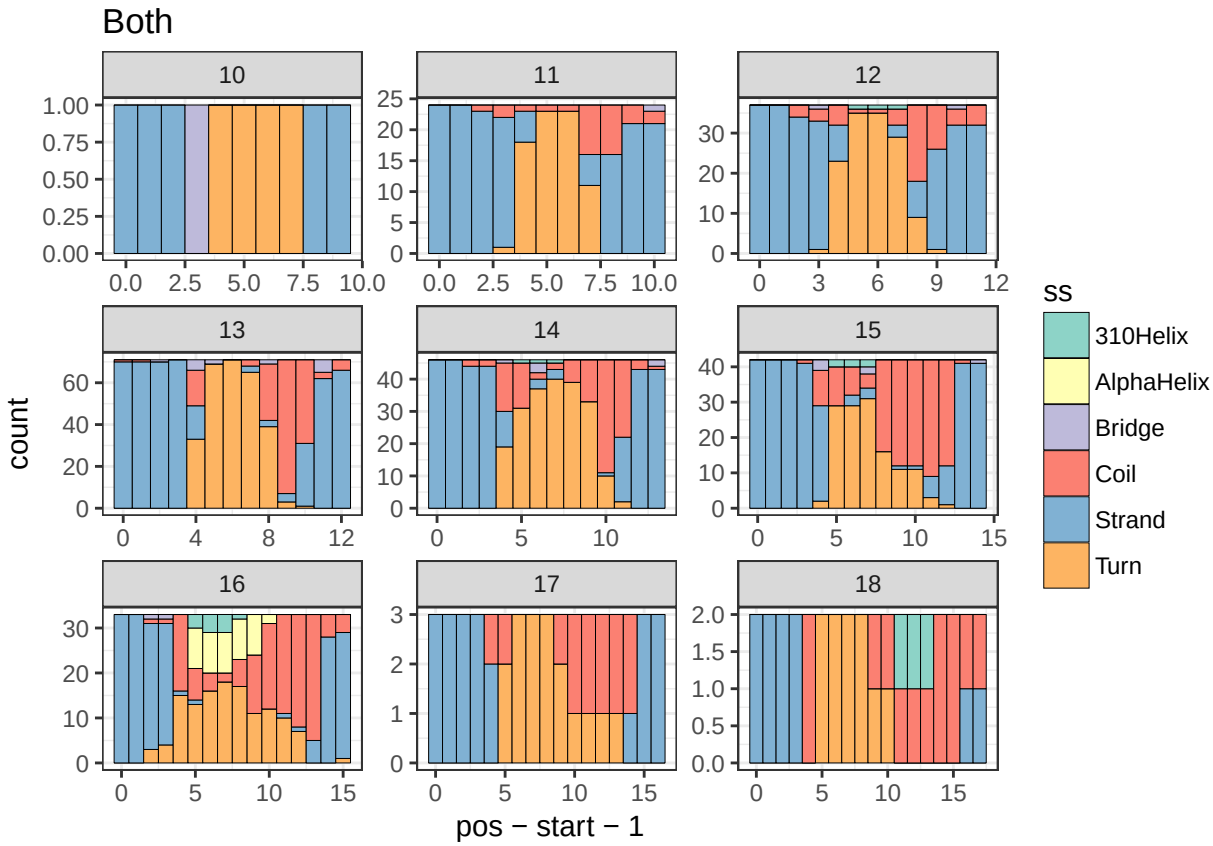
```
  facet_wrap(~len_tcr, scales = "free") +
  scale_fill_brewer(palette = "Set3", drop=F) +
  theme_bw() + ggtitle("TRA")
```




```
ggplot(df.ss.1 %>% filter(region=="CDR3" & len_tcr > 5 & tcr_gene=="TRB"), aes(x=pos - start - 1, fill =
  geom_histogram(binwidth = 1, color = "black", size = 0.1) +
  facet_wrap(~len_tcr, scales = "free") +
  scale_fill_brewer(palette = "Set3", drop=F) +
  theme_bw() + ggtitle("TRB")
```

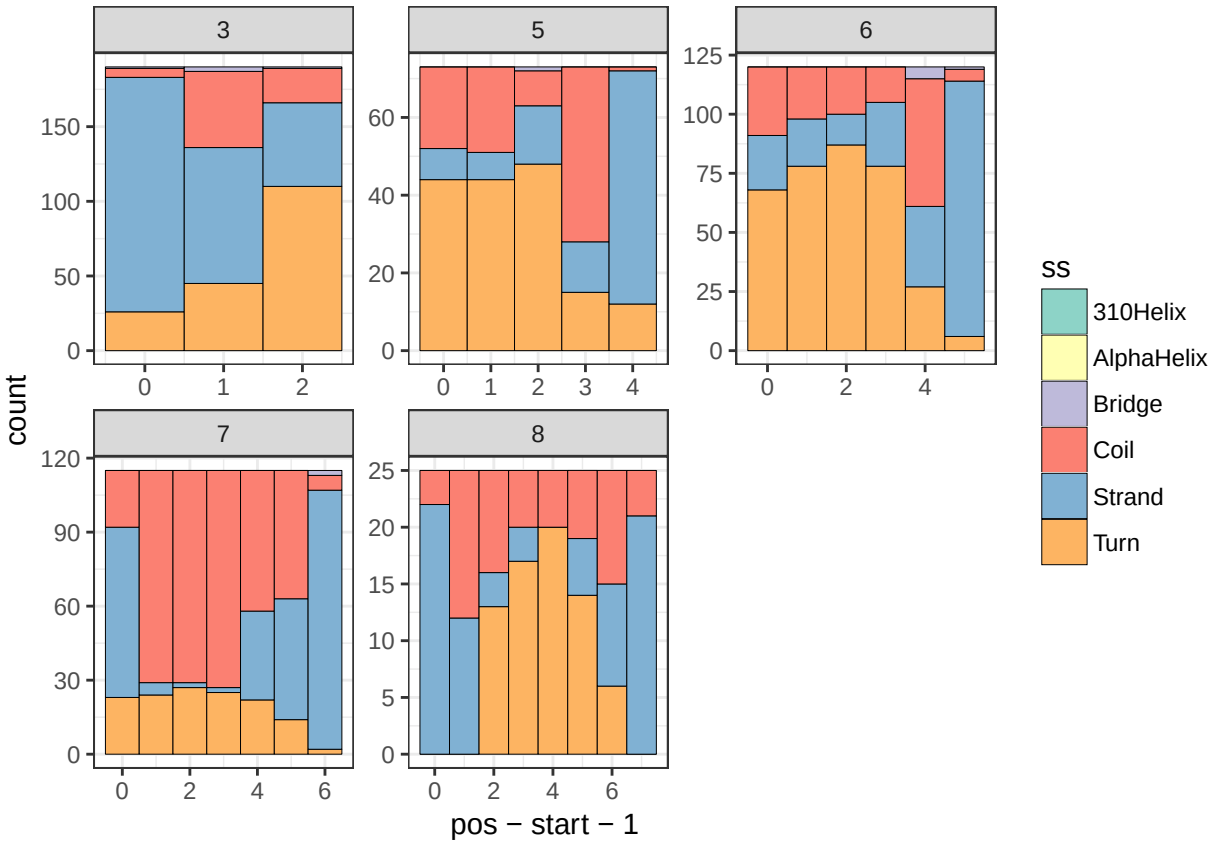



```
ggplot(df.ss.1 %>% filter(region=="CDR3" & len_tcr > 5), aes(x=pos - start - 1, fill = ss)) +
  geom_histogram(binwidth = 1, color = "black", size = 0.1) +
  facet_wrap(~len_tcr, scales = "free") +
  scale_fill_brewer(palette = "Set3", drop=F) +
  theme_bw() + ggtitle("Both")
```



CDR1-2 regions

```
ggplot(df.ss.1 %>% filter(region!="CDR3"), aes(x=pos - start - 1, fill = ss)) +
  geom_histogram(binwidth = 1, color = "black", size = 0.1) +
  facet_wrap(~len_tcr, scales = "free") +
  scale_fill_brewer(palette = "Set3", drop = F) +
  theme_bw()
```


3
5
6
7
8
count
pos − start − 1
SS
310Helix
AlphaHelix
Bridge
Coil
Strand
Turn