

Data Analysis

2022-11-09

Read in data

In the first step the dataset is read in and the column "ASAFind 25aa sequence" is split into a new column after each value. This way we can analyze the first 15 combinations.

```
data <- read_excel("/Users/fuadsarkerjeem/Desktop/Final Datasets/11.11.22/All_together.xlsx")

## New names:
## * '' -> '...1'
## * 'Stability ASAP(25aa)' -> 'Stability ASAP(25aa)...31'
## * 'Stability ASAP(25aa)' -> 'Stability ASAP(25aa)...33'

# Overview of the data set
```

Comparison for Plastid only

First we select the relevant columns (charges and Location) and filter the data for Plastid and then pivot from wide to long before plotting the boxplot. The red dot represents the mean as addition to the median.

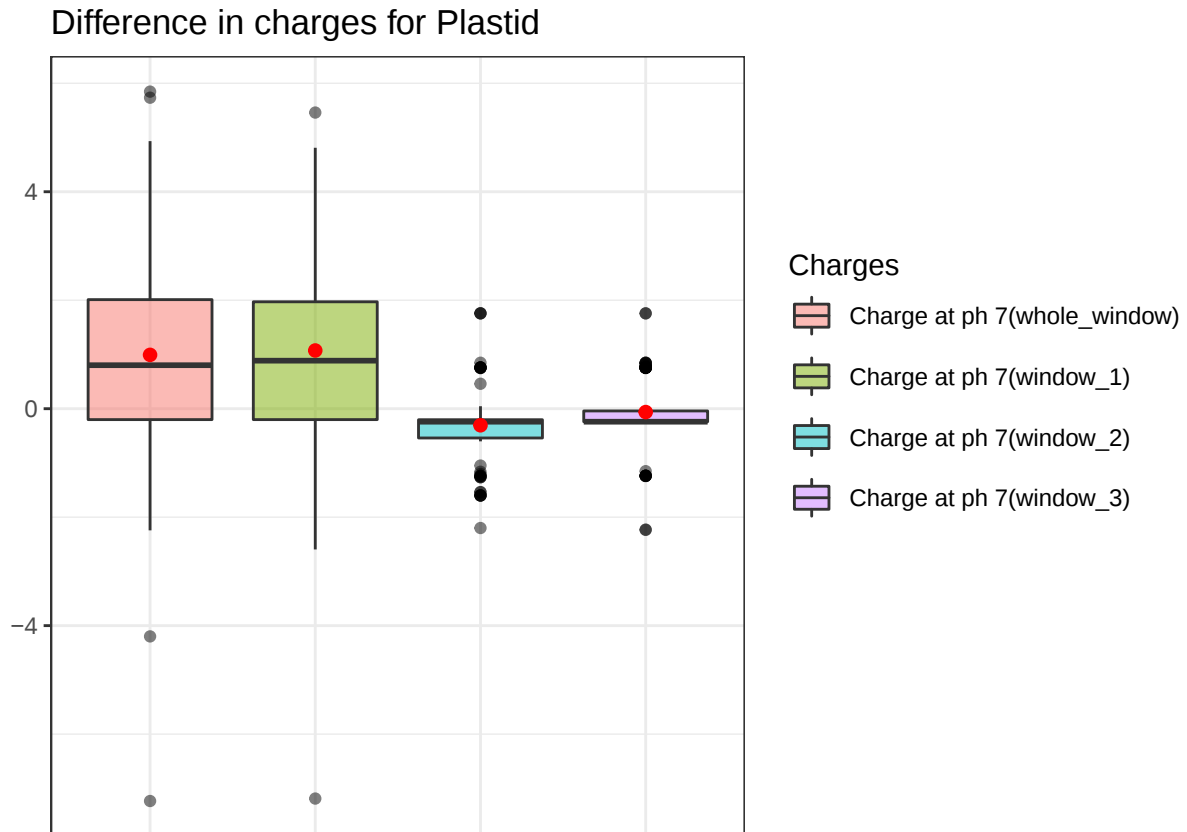
```
data_comp <- data %>%
  dplyr::select(Location, `Charge at ph 7(whole_window)`,
                `Charge at ph 7(window_2)`,
                `Charge at ph 7(window_1)`,
                `Charge at ph 7(window_3)`)

data_plast <- data_comp %>%
  filter(Location == "Plastid")

# pivot from wide to long
data_plast2 <- data_plast %>%
  pivot_longer(!Location, names_to = "Charges", values_to = "value")

# boxplot
data_plast2 %>%
  ggplot(aes(x = Charges, y = value, fill = Charges)) +
  geom_boxplot(outlier.colour="black", alpha = 0.5) +
  stat_summary(fun=mean, geom="point", shape=20, size=3, color="red", fill="red") +
  labs(title="Difference in charges for Plastid",
        x="", y = "") +
  theme_bw() +
  theme(legend.position = "right",
```

```
axis.title.x=element_blank(),
axis.text.x=element_blank(),
axis.ticks.x=element_blank()) +
guides(fill=guide_legend(nrow=6,byrow=TRUE, reverse=FALSE))
```



Anova for Plastids difference in charge

We fit an ANOVA model to see if there are significant differences between the charge variables for Plastid:

```
# ANOVA model
plastid_aov <- aov(value ~ Charges,
  data = data_plast2
)

# summary
summary(plastid_aov) # we can see from summary that there are significant differences
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Charges      3   204.3    68.11   35.99 <2e-16 ***
## Residuals   536  1014.3     1.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# display it as a data frame and perform post hoc analysis
```

```
plastid_aov <- data_plast2 %>%  
  tukey_hsd(value ~ Charges)
```

```
plastid_aov
```

```
## # A tibble: 6 x 9
```

```
##   term      group1      group2    null.value estimate conf.low conf.high   p.adj  
## * <chr>   <chr>      <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>  
## 1 Charges Charge at ~ Charge at~      0    0.0806   -0.351    0.512 9.63e- 1  
## 2 Charges Charge at ~ Charge at~      0   -1.30     -1.73    -0.867 3.37e-10  
## 3 Charges Charge at ~ Charge at~      0   -1.05     -1.49    -0.623 4.1 e- 9  
## 4 Charges Charge at ~ Charge at~      0   -1.38     -1.81    -0.948 3.37e-10  
## 5 Charges Charge at ~ Charge at~      0   -1.14     -1.57    -0.704 5.28e-10  
## 6 Charges Charge at ~ Charge at~      0    0.244    -0.188    0.675 4.65e- 1  
## # ... with 1 more variable: p.adj.signif <chr>
```

From the ANOVA model we can see that the difference between:

- Charge at ph 7(ASAP 25aa)-Charge at ph 7(+1 poition to 20th)
- Charge at ph 7(charges_window_small)-Charge at ph 7(charges_window_2)

are insignificant. Meanwhile the differences between:

- Charge at ph 7(charges_window_2)-Charge at ph 7(+1 poition to 20th)
– -1.379
- Charge at ph 7(charges_window_small)-Charge at ph 7(+1 poition to 20th)
– -1.135
- Charge at ph 7(charges_window_2)-Charge at ph 7(ASAP 25aa)
– -1.298
- Charge at ph 7(charges_window_2)-Charge at ph 7(ASAP 25aa)
– -1.054

Are statistically significant with p-values < 0.05 (can be seen in the above table).

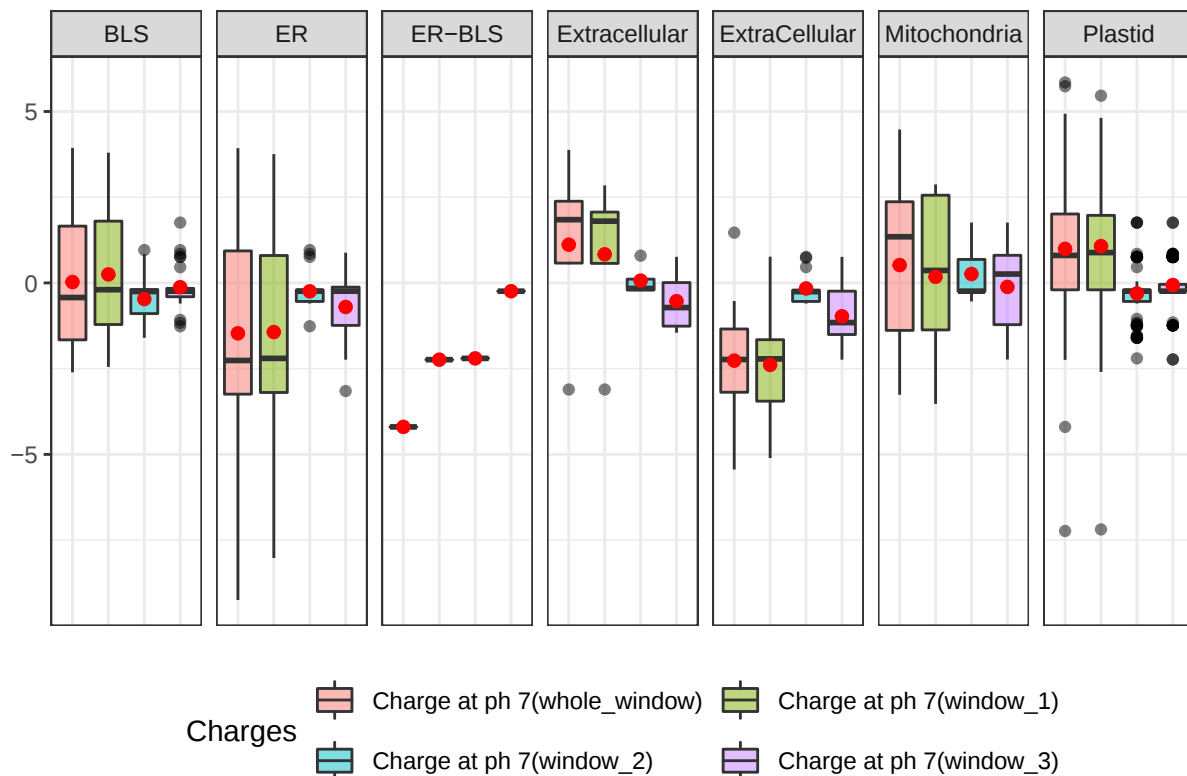
Comparison between all Locations

First we create boxplots to visualize range and comparison between charges and groups. For this we flip the data from wide to long format:

```
# wide to long
data_comp2 <- data_comp %>%
  pivot_longer(!Location, names_to = "Charges", values_to = "value")

# Version 1 facettted by Location
data_comp2 %>%
  ggplot(aes(x = Charges, y = value, fill = Charges)) +
  geom_boxplot(outlier.colour="black", alpha = 0.5) +
  stat_summary(fun=mean, geom="point", shape=20, size=3, color="red", fill="red") +
  facet_grid(~ Location) +
  labs(title="Difference in charges across locations",
       x=" ", y = "") +
  theme_bw() +
  theme(legend.position = "bottom",
        axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) +
  guides(fill=guide_legend(nrow=2,byrow=TRUE, reverse=FALSE))
```

Difference in charges across locations



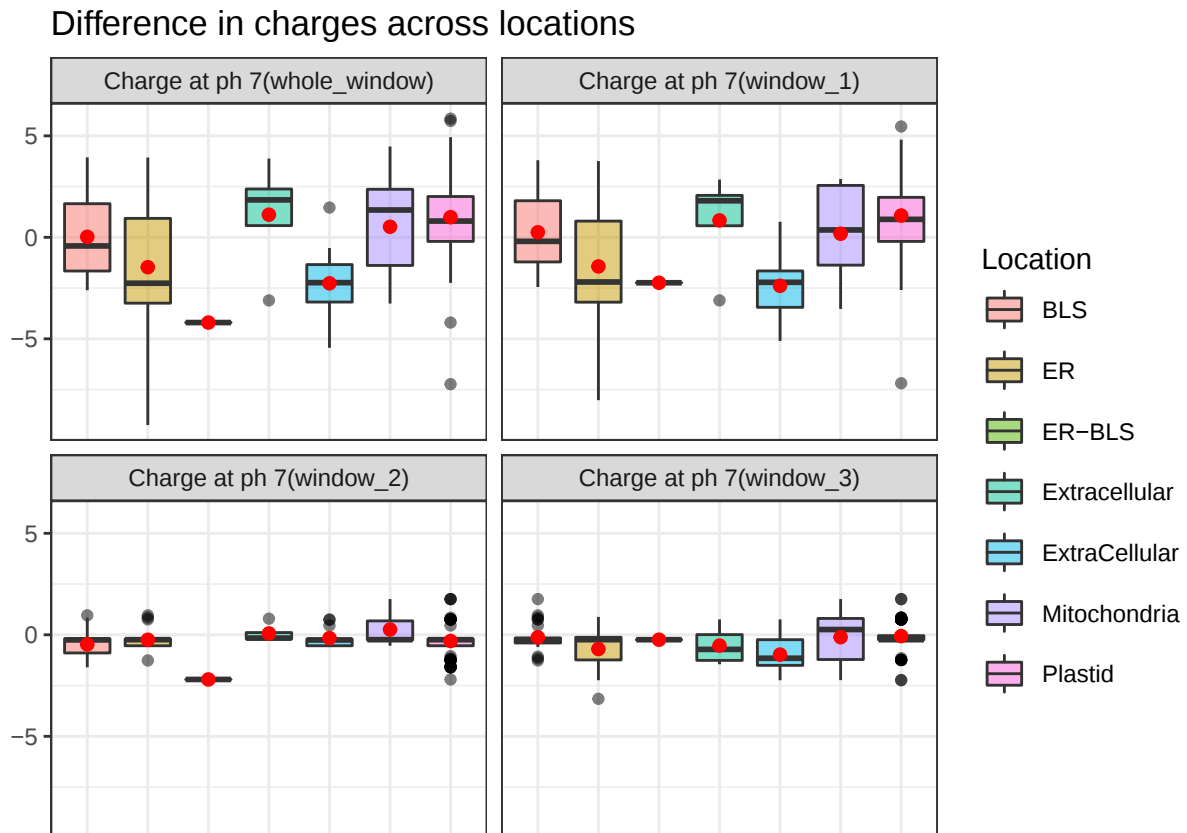
Second version were we group by charge:

```
# Version 1 faceted by charge
data_comp2 %>%
  ggplot(aes(x = Location, y = value, fill = Location)) +
  geom_boxplot(outlier.colour="black", alpha = 0.5) +
  stat_summary(fun=mean, geom="point", shape=20, size=3, color="red", fill="red") + # add mean
```

```

facet_wrap(~ Charges, nrow = 2) + # facet
labs(title="Difference in charges across locations",
      x=" ", y = "") +
theme_bw() +
theme(legend.position = "right",
      axis.title.x=element_blank(),
      axis.text.x=element_blank(),
      axis.ticks.x=element_blank()) +
guides(fill=guide_legend(nrow=7,byrow=TRUE, reverse=FALSE))

```



Anova model for all Locations

```

# ANOVA model
all_aov <- aov(value ~ Charges + Location,
               data = data_comp2
)

# summary
summary(all_aov) # we can see from summary that there are significant differences

```

```

##           Df Sum Sq Mean Sq F value    Pr(>F)
## Charges      3   79.1    26.35   10.00 1.74e-06 ***

```

```
## Location      6  337.0   56.17   21.32  < 2e-16 ***
## Residuals    862 2270.4    2.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# display it as a data frame and perform post hoc analysis
```

```
all_aov <- data_comp2 %>%
  tukey_hsd(value ~ Charges * Location)
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 108 rows [8, 13, 18,
## 19, 20, 21, 35, 36, 37, 38, 61, 62, 63, 64, 86, 87, 88, 89, 110, 111, ...].
```

```
all_aov
```

```
## # A tibble: 405 x 9
##   term      group1      group2 null.value estimate conf.low conf.high  p.adj
##   * <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 Charges Charge at ~ Charge a~      0  0.0622   -0.322    0.446 9.76e-1
## 2 Charges Charge at ~ Charge a~      0 -0.601   -0.984   -0.217 3.56e-4
## 3 Charges Charge at ~ Charge a~      0 -0.535   -0.919   -0.151 2 e-3
## 4 Charges Charge at ~ Charge a~      0 -0.663   -1.05    -0.279 5.86e-5
## 5 Charges Charge at ~ Charge a~      0 -0.597   -0.981   -0.213 3.92e-4
## 6 Charges Charge at ~ Charge a~      0  0.0657   -0.318    0.449 9.71e-1
## 7 Location BLS      ER      0 -0.882   -1.51    -0.256 6.8 e-4
## 8 Location BLS      ER      0 -2.14    -4.48     0.203 9.96e-2
## 9 Location BLS      Extracel~      0  0.451   -0.781     1.68 9.33e-1
## 10 Location BLS      ExtraCel~      0 -1.37    -2.13    -0.610 2.54e-6
## # ... with 395 more rows, and 1 more variable: p.adj.signif <chr>
```

This is an ANOVA model that gives you the combination of Charges * Location, so we are looking at the interplay of these two. You can see for example:

- Charge at ph 7(+1 poition to 20th):BLS - Charge at ph 7(+1 poition to 20th):ExtraCellular
- Charge at ph 7(+1 poition to 20th):BLS - Charge at ph 7(ASAP 25aa):ExtraCellular
- Charge at ph 7(ASAP 25aa):ExtraCellular - Charge at ph 7(ASAP 25aa):Plastid

and so on would be significant. We won't analyze all the rows here since the amount of results is quite large with 405 different combinations and depending on individual preference results can be extracted from the table. However since a lot of the results are not significant, in the next code line we remove all not significant ones:

```
# filter out p.adj.signif = ns
all_aov_signif <- all_aov %>%
  filter(p.adj.signif != "ns")
```

```
all_aov_signif
```

```
## # A tibble: 62 x 9
##   term      group1      group2 null.value estimate conf.low conf.high  p.adj
##   <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
```

```
## 1 Charges Charge at~ Charge a~      0 -0.601 -0.984      -0.217 3.56e- 4
## 2 Charges Charge at~ Charge a~      0 -0.535 -0.919      -0.151 2 e- 3
## 3 Charges Charge at~ Charge a~      0 -0.663 -1.05       -0.279 5.86e- 5
## 4 Charges Charge at~ Charge a~      0 -0.597 -0.981      -0.213 3.92e- 4
## 5 Location BLS      ER      0 -0.882 -1.51       -0.256 6.8 e- 4
## 6 Location BLS      ExtraCel~      0 -1.37 -2.13       -0.610 2.54e- 6
## 7 Location BLS      Plastid      0 0.504 0.0195      0.989 3.53e- 2
## 8 Location ER      Extracel~      0 1.33 0.101       2.57 2.41e- 2
## 9 Location ER      Mitochon~      0 1.17 0.321       2.02 1 e- 3
## 10 Location ER      Plastid      0 1.39 0.902       1.87 1.49e-13
## # ... with 52 more rows, and 1 more variable: p.adj.signif <chr>
```

The results show that there are a total of 62 combinations that are significant from the ANOVA model.

Amino sequence

To see the position of each letter we create a new column (x1, x2 , x3 ...) and simply split the column off at the first location (first letter), the second location etc. so we are left with the single Letter from each position in a new column.

```
# glimpse(data)

data1 <- data

# create new columns with each respective letter from position 1, 2, 3 and so on ...
data1 <- data1 %>%
  mutate(x1 = substr(data1$`ASAFind 25aa sequence`, start = 1, stop = 1),
         x2 = substr(data1$`ASAFind 25aa sequence`, start = 2, stop = 2),
         x3 = substr(data1$`ASAFind 25aa sequence`, start = 3, stop = 3),
         x4 = substr(data1$`ASAFind 25aa sequence`, start = 4, stop = 4),
         x5 = substr(data1$`ASAFind 25aa sequence`, start = 5, stop = 5),
         x6 = substr(data1$`ASAFind 25aa sequence`, start = 6, stop = 6),
         x7 = substr(data1$`ASAFind 25aa sequence`, start = 7, stop = 7),
         x8 = substr(data1$`ASAFind 25aa sequence`, start = 8, stop = 8),
         x9 = substr(data1$`ASAFind 25aa sequence`, start = 9, stop = 9),
         x10 = substr(data1$`ASAFind 25aa sequence`, start = 10, stop = 10),
         x11 = substr(data1$`ASAFind 25aa sequence`, start = 11, stop = 11),
         x12 = substr(data1$`ASAFind 25aa sequence`, start = 12, stop = 12),
         x13 = substr(data1$`ASAFind 25aa sequence`, start = 13, stop = 13),
         x14 = substr(data1$`ASAFind 25aa sequence`, start = 14, stop = 14),
         x15 = substr(data1$`ASAFind 25aa sequence`, start = 15, stop = 15),
         x16 = substr(data1$`ASAFind 25aa sequence`, start = 16, stop = 16),
         x17 = substr(data1$`ASAFind 25aa sequence`, start = 17, stop = 17),
         x18 = substr(data1$`ASAFind 25aa sequence`, start = 18, stop = 18),
         x19 = substr(data1$`ASAFind 25aa sequence`, start = 19, stop = 19),
         x20 = substr(data1$`ASAFind 25aa sequence`, start = 20, stop = 20),
         x21 = substr(data1$`ASAFind 25aa sequence`, start = 21, stop = 21),
         x22 = substr(data1$`ASAFind 25aa sequence`, start = 22, stop = 22),
         x23 = substr(data1$`ASAFind 25aa sequence`, start = 23, stop = 23),

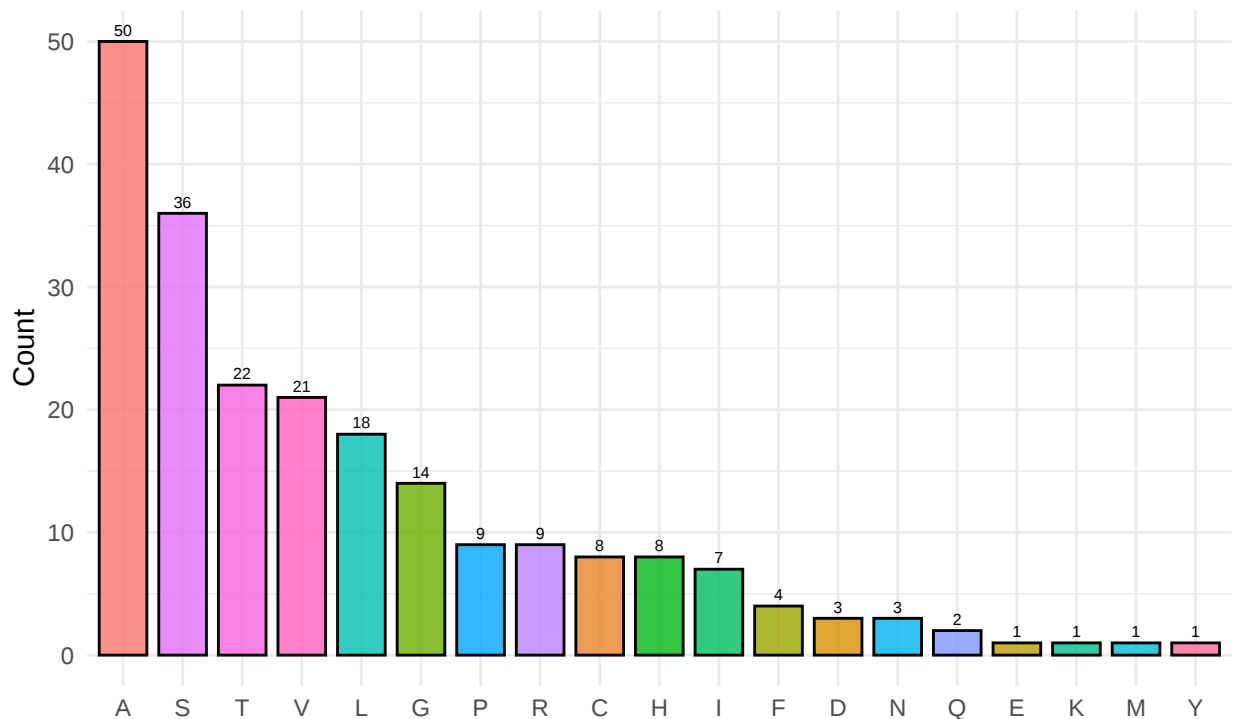
  )
```

Last step we create a count of the letters and visualize it in a descending bar chart for each new column.

Pos 1

```
data1 %>%
  dplyr::count(x1) %>%
  ggplot(aes(x = reorder(x1, desc(n)), y = n, fill = x1)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 1",
        subtitle = "",
        x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

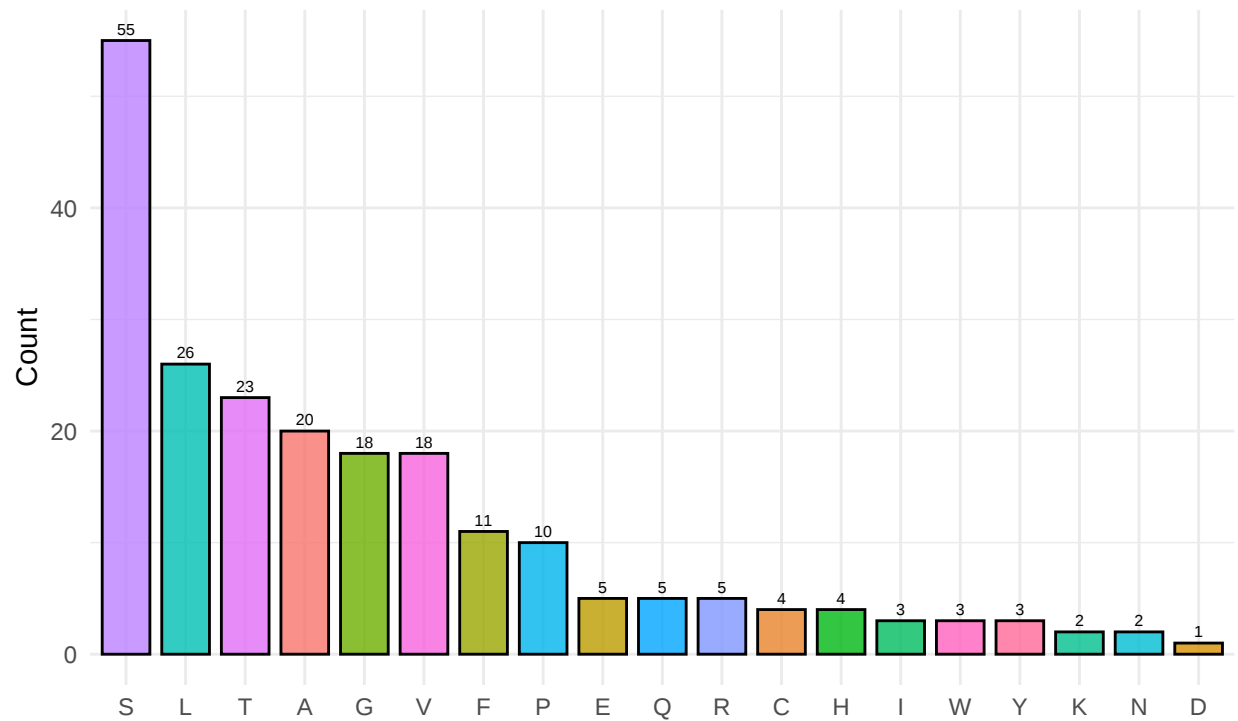
Amoniacid count Position 1



Pos 2

```
data1 %>%
  dplyr::count(x2) %>%
  ggplot(aes(x = reorder(x2, desc(n)), y = n, fill = x2)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 2",
        subtitle = "",
        x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

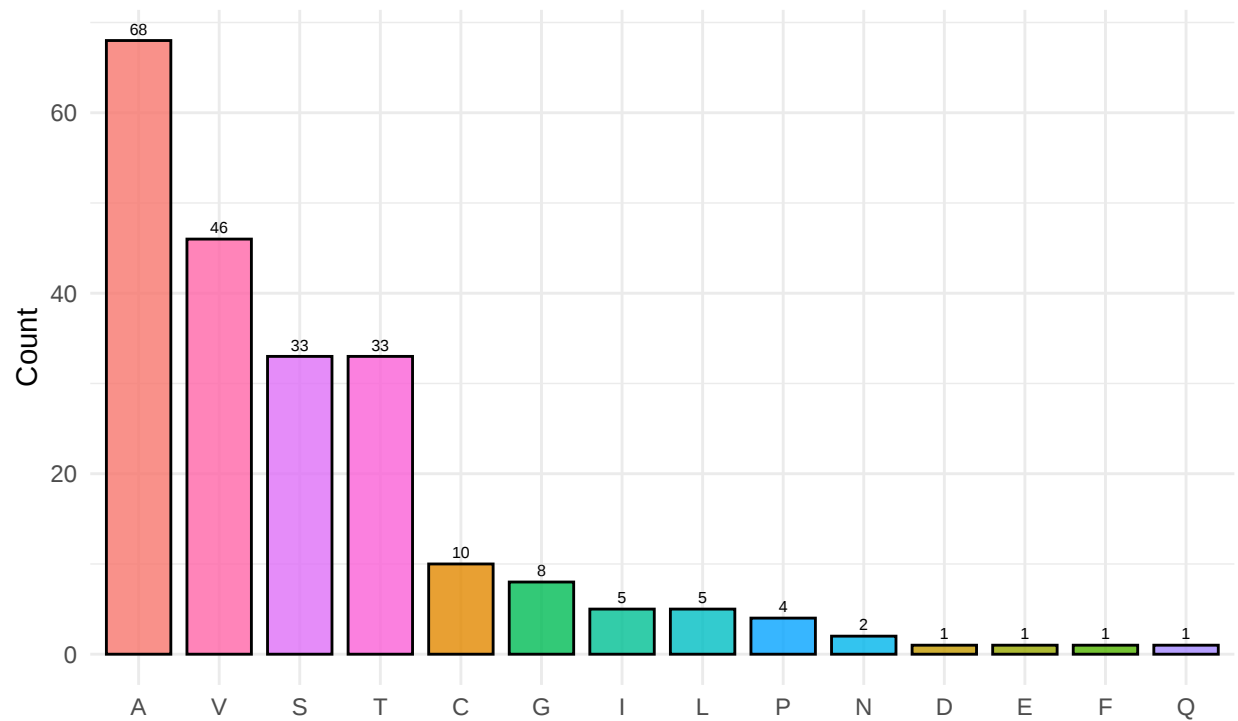
Amoniacid count Position 2



Pos 3

```
data1 %>%
  dplyr::count(x3) %>%
  ggplot(aes(x = reorder(x3, desc(n)), y = n, fill = x3)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 3",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

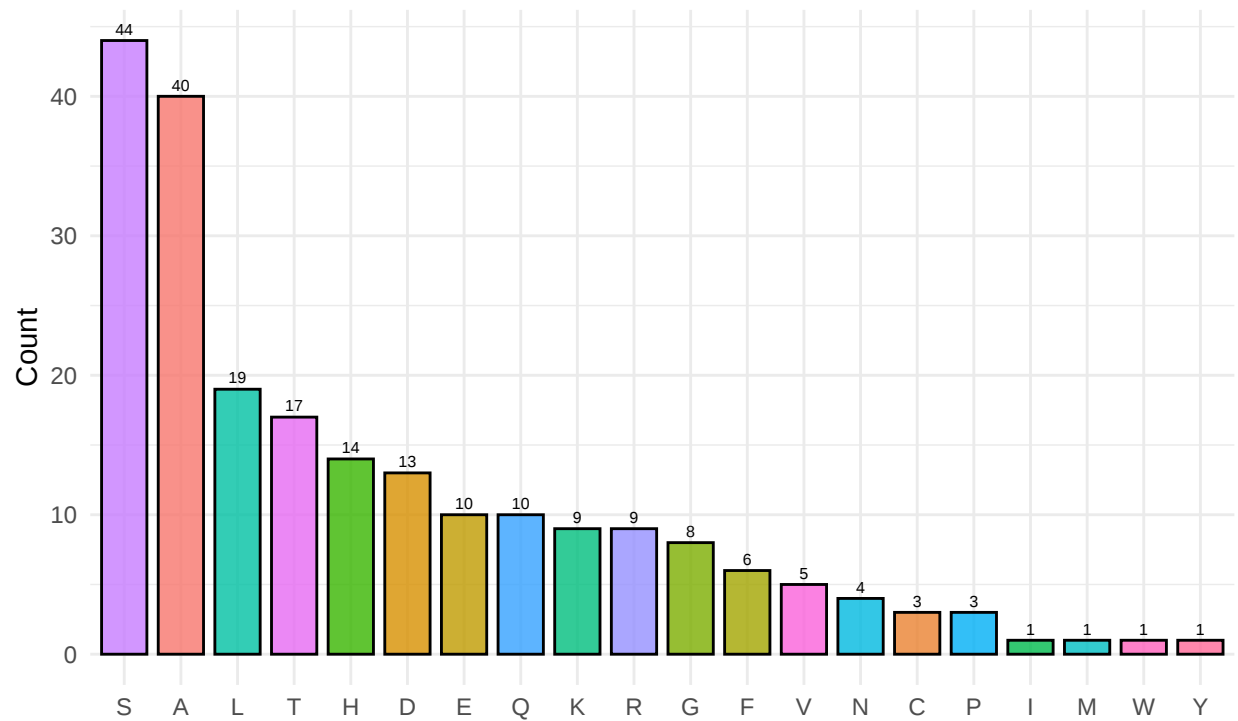
Amoniacid count Position 3



Pos 4

```
data1 %>%
  dplyr::count(x4) %>%
  ggplot(aes(x = reorder(x4, desc(n)), y = n, fill = x4)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 4",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

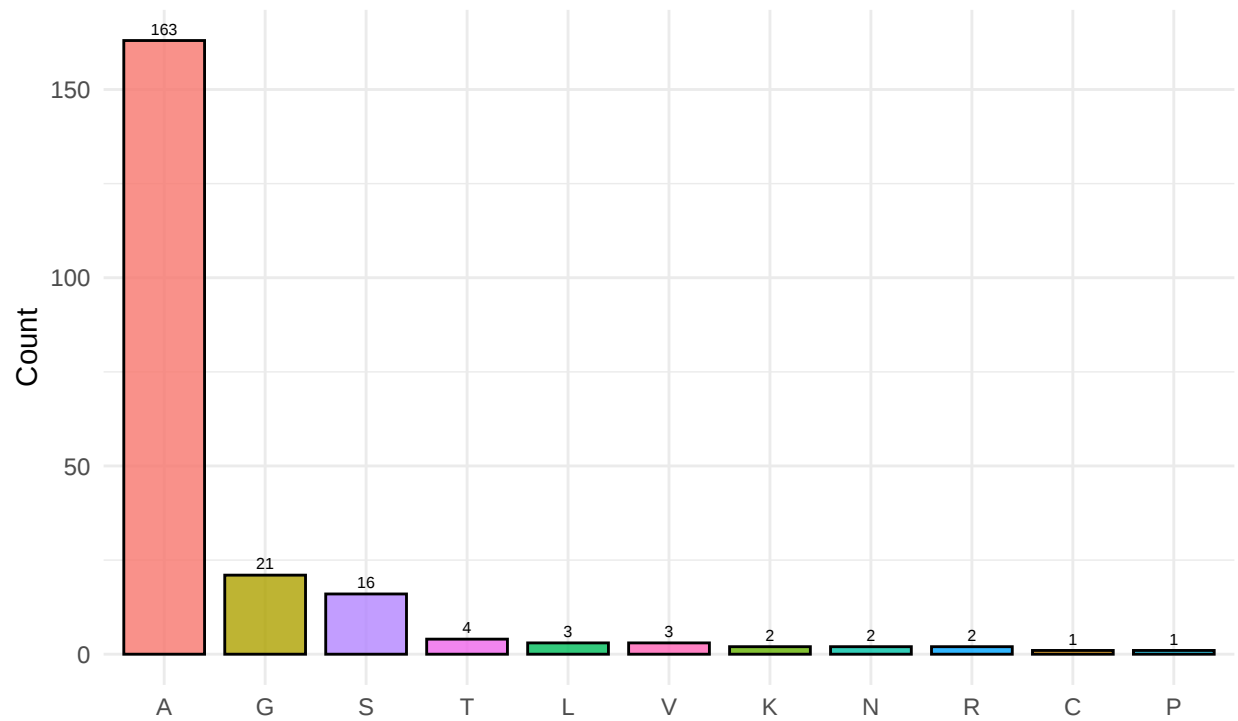
Amoniacid count Position 4



Pos 5

```
data1 %>%
  dplyr::count(x5) %>%
  ggplot(aes(x = reorder(x5, desc(n)), y = n, fill = x5)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 5",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

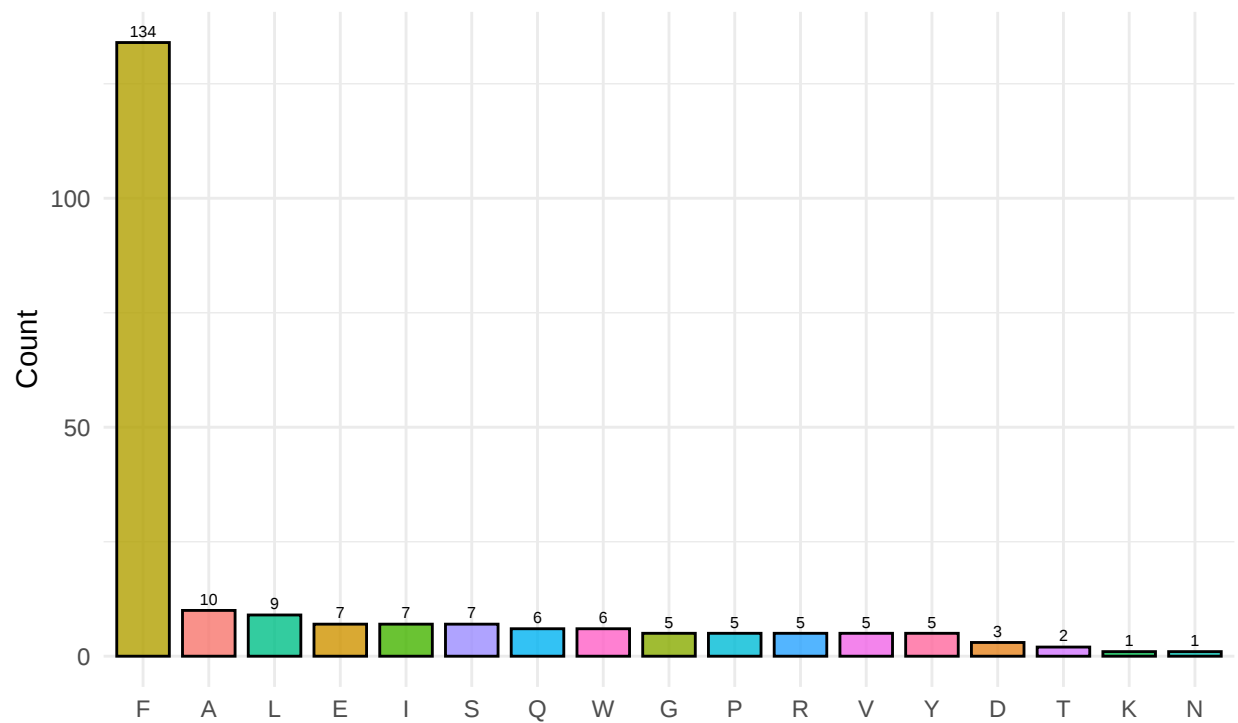
Amoniacid count Position 5



Pos 6

```
data1 %>%
  dplyr::count(x6) %>%
  ggplot(aes(x = reorder(x6, desc(n)), y = n, fill = x6)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 6",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

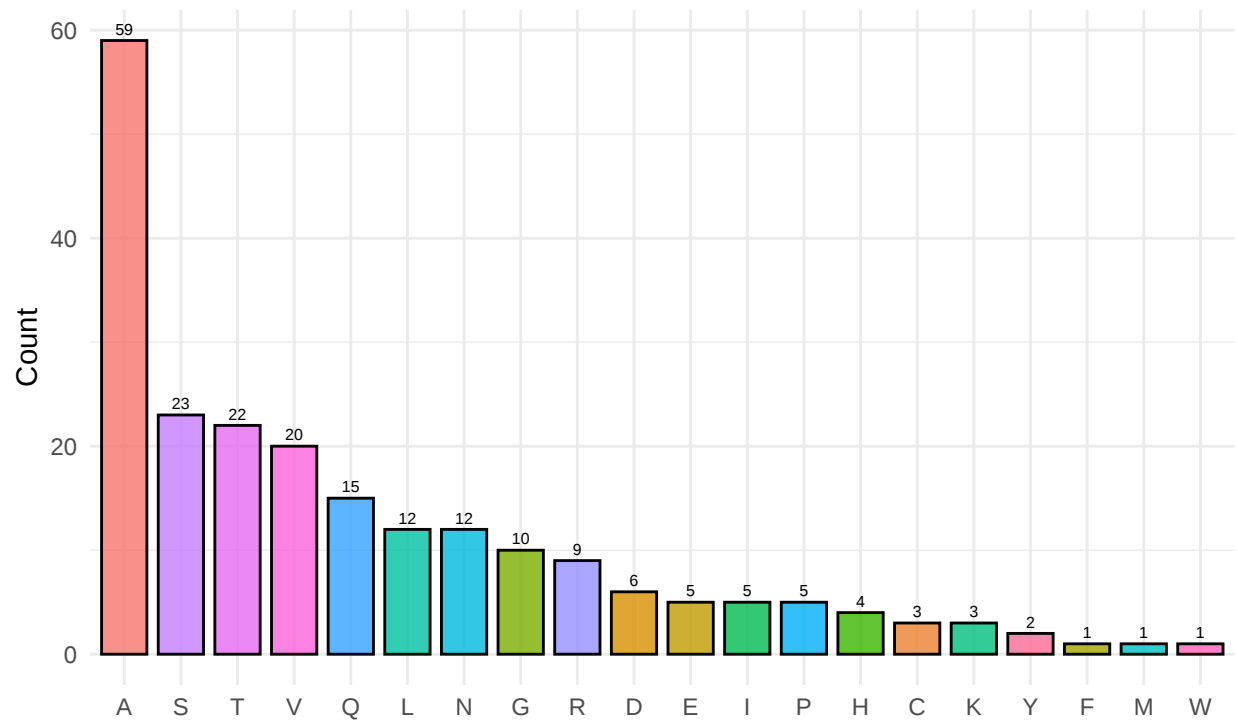
Amoniacid count Position 6



Pos 7

```
data1 %>%
  dplyr::count(x7) %>%
  ggplot(aes(x = reorder(x7, desc(n)), y = n, fill = x7)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 7",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

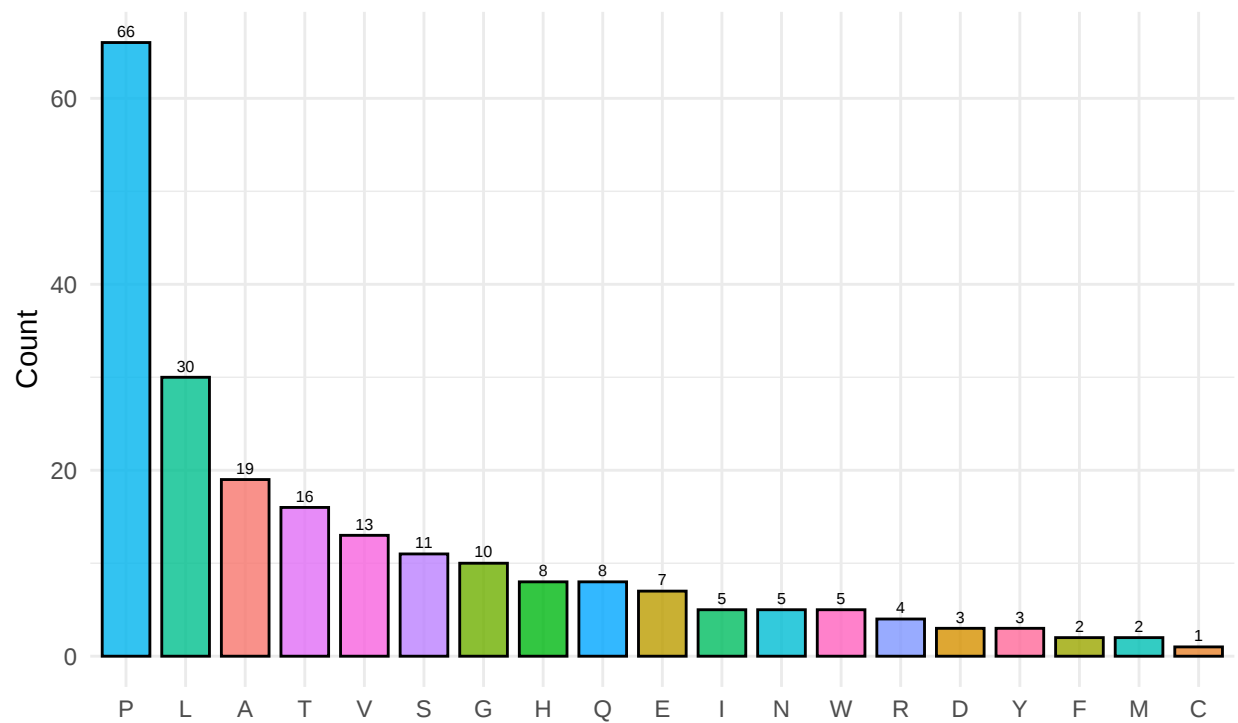
Amoniacid count Position 7



Pos 8

```
data1 %>%
  dplyr::count(x8) %>%
  ggplot(aes(x = reorder(x8, desc(n)), y = n, fill = x8)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 8",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

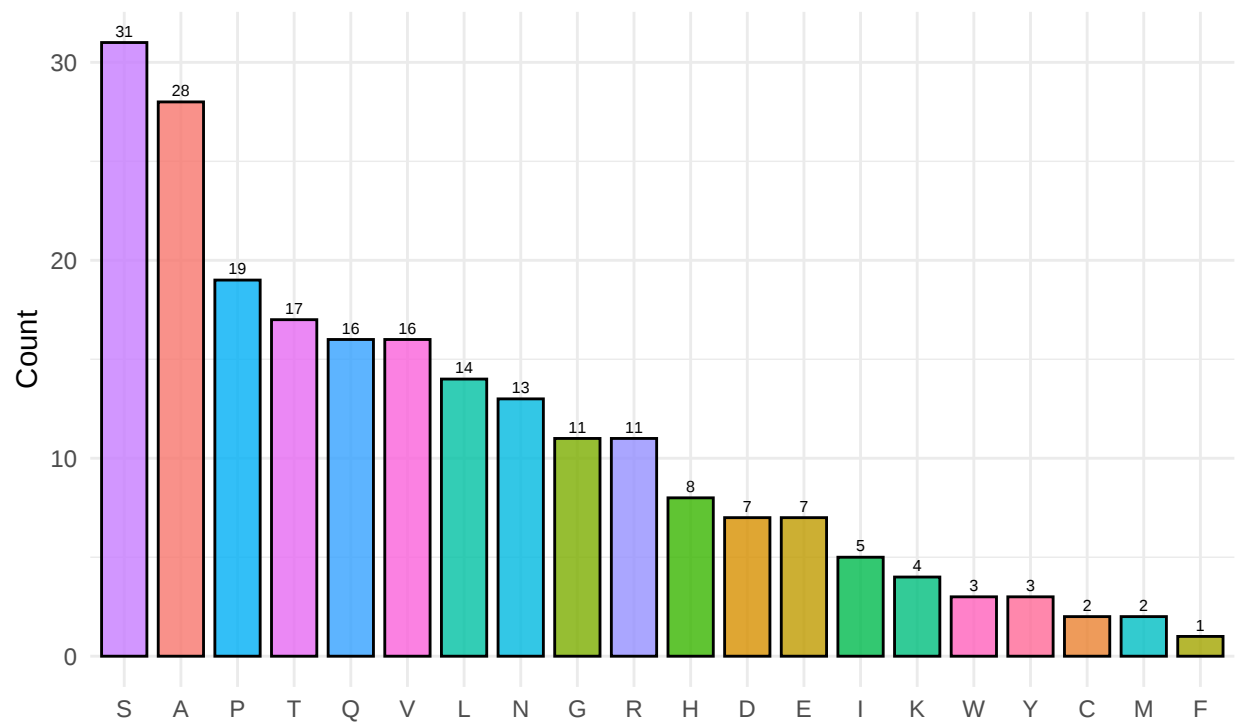
Amoniacid count Position 8



Pos 9

```
data1 %>%
  dplyr::count(x9) %>%
  ggplot(aes(x = reorder(x9, desc(n)), y = n, fill = x9)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 9",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

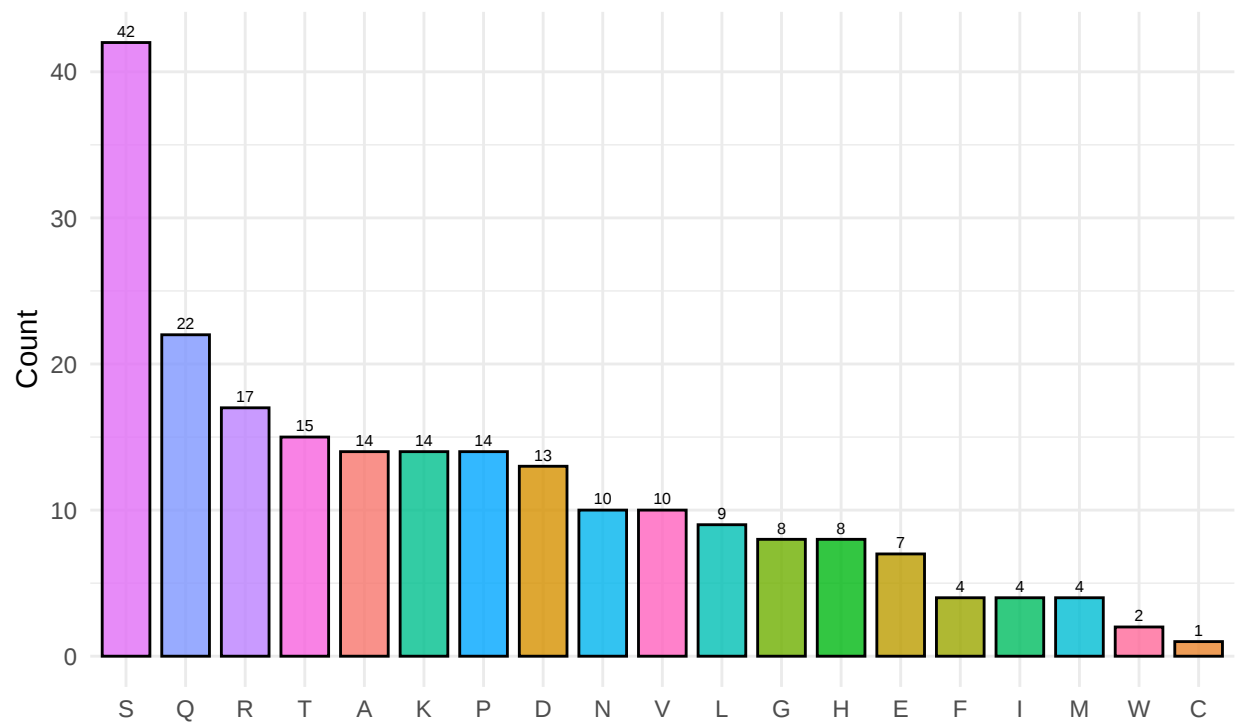
Amoniacid count Position 9



Pos 10

```
data1 %>%
  dplyr::count(x10) %>%
  ggplot(aes(x = reorder(x10, desc(n)), y = n, fill = x10)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 10",
        subtitle = "",
        x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

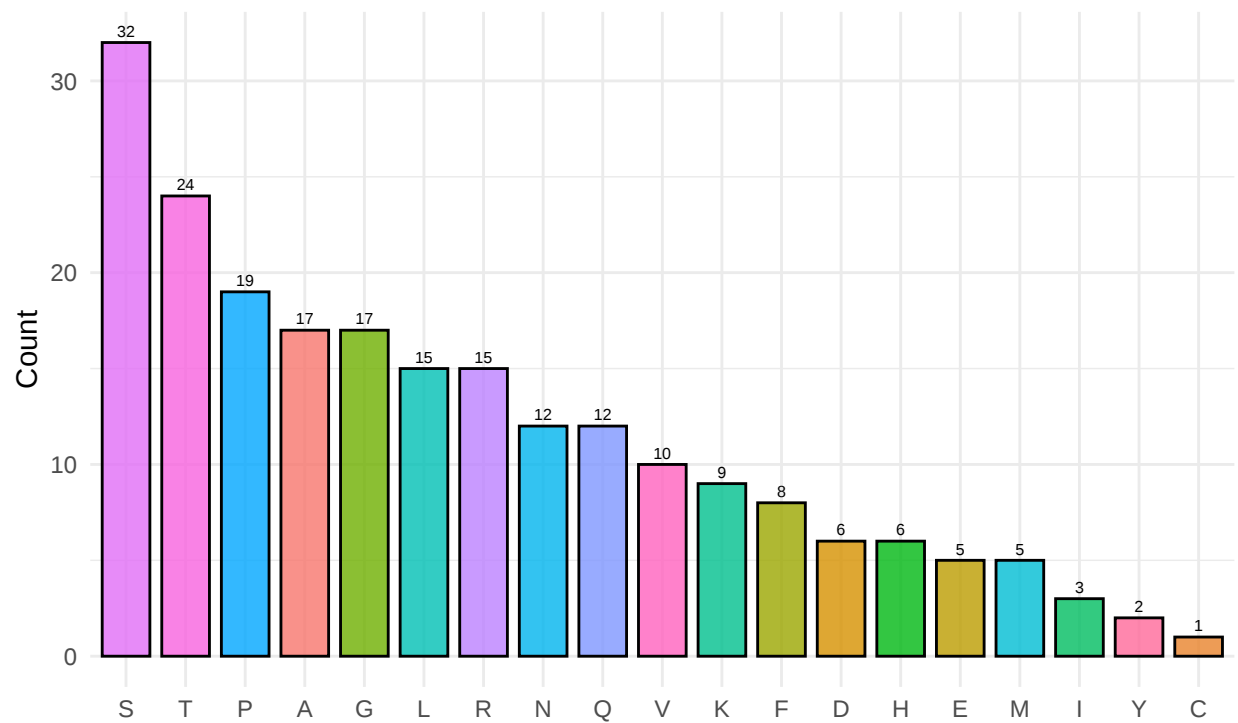
Amoniacid count Position 10



Pos 11

```
data1 %>%
  dplyr::count(x11) %>%
  ggplot(aes(x = reorder(x11, desc(n)), y = n, fill = x11)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n, position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
    labs(title="Amoniacid count Position 11",
      subtitle = "",
      x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

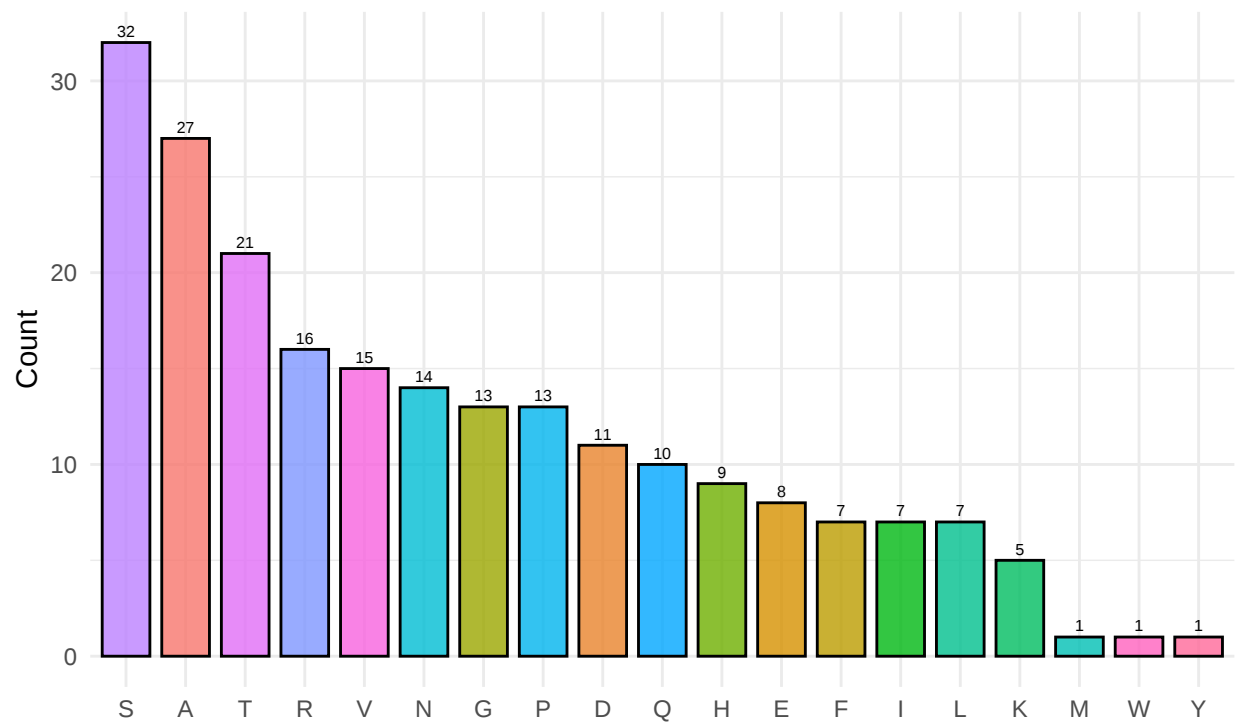
Amoniacid count Position 11



Pos 12

```
data1 %>%
  dplyr::count(x12) %>%
  ggplot(aes(x = reorder(x12, desc(n)), y = n, fill = x12)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 12",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

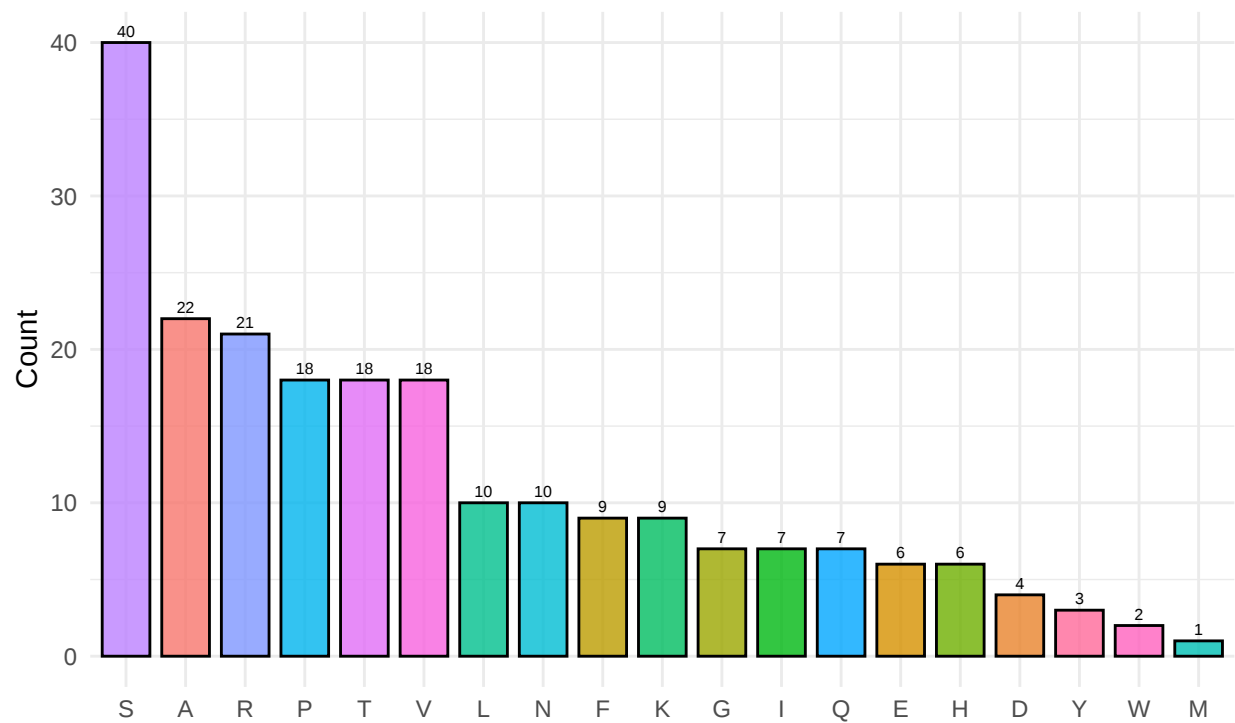
Amoniacid count Position 12



Pos 13

```
data1 %>%
  dplyr::count(x13) %>%
  ggplot(aes(x = reorder(x13, desc(n)), y = n, fill = x13)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 13",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

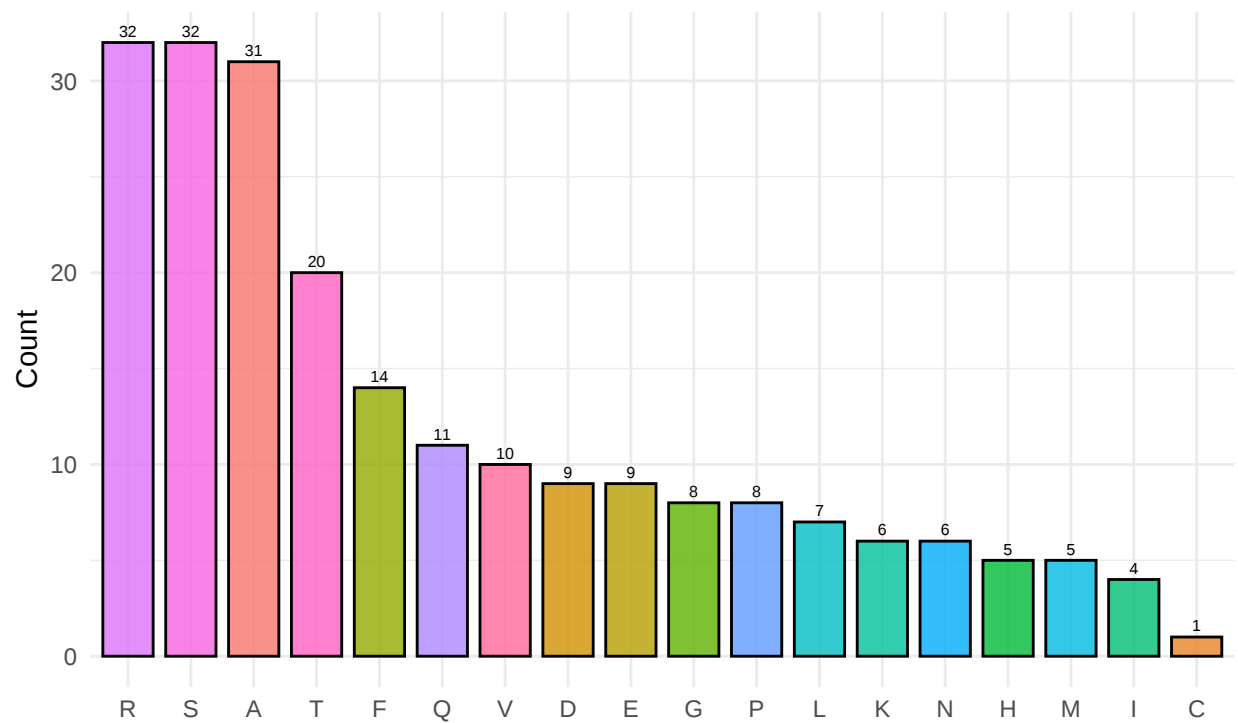
Amoniacid count Position 13



Pos 14

```
data1 %>%
  dplyr::count(x14) %>%
  ggplot(aes(x = reorder(x14, desc(n)), y = n, fill = x14)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 14",
        subtitle = "",
        x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

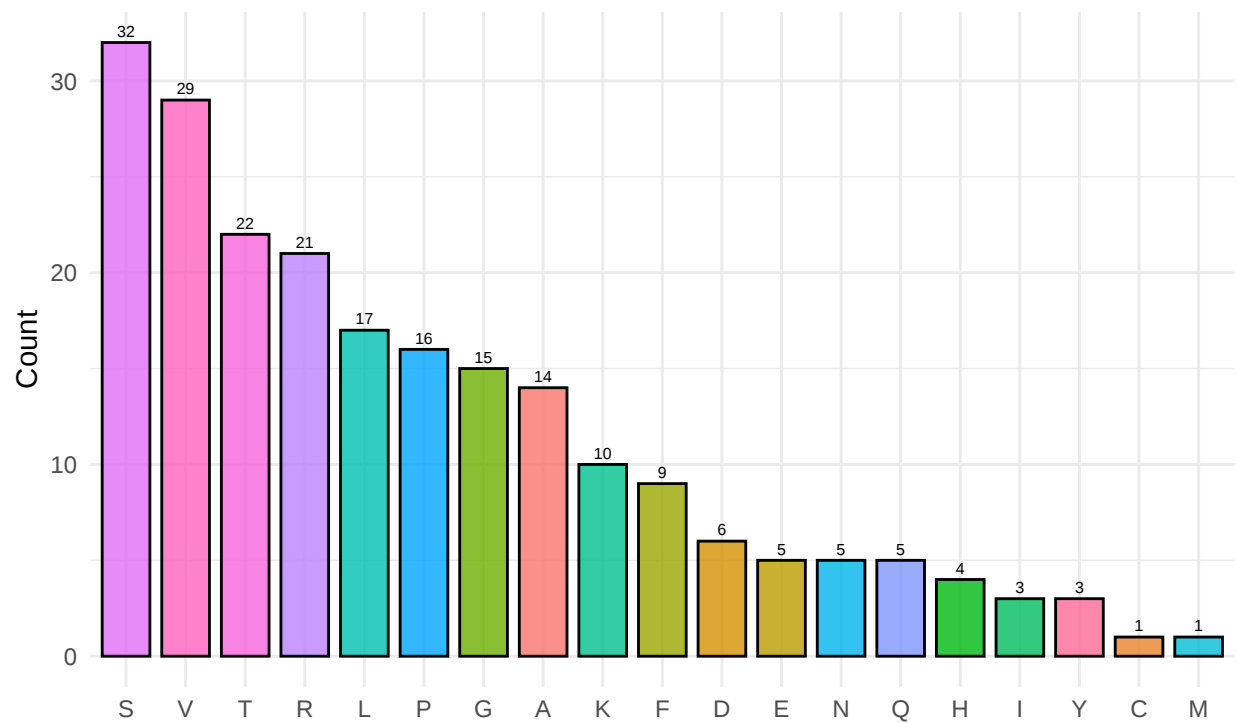
Amoniacid count Position 14



Pos 15

```
data1 %>%
  dplyr::count(x15) %>%
  ggplot(aes(x = reorder(x15, desc(n)), y = n, fill = x15)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 15",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

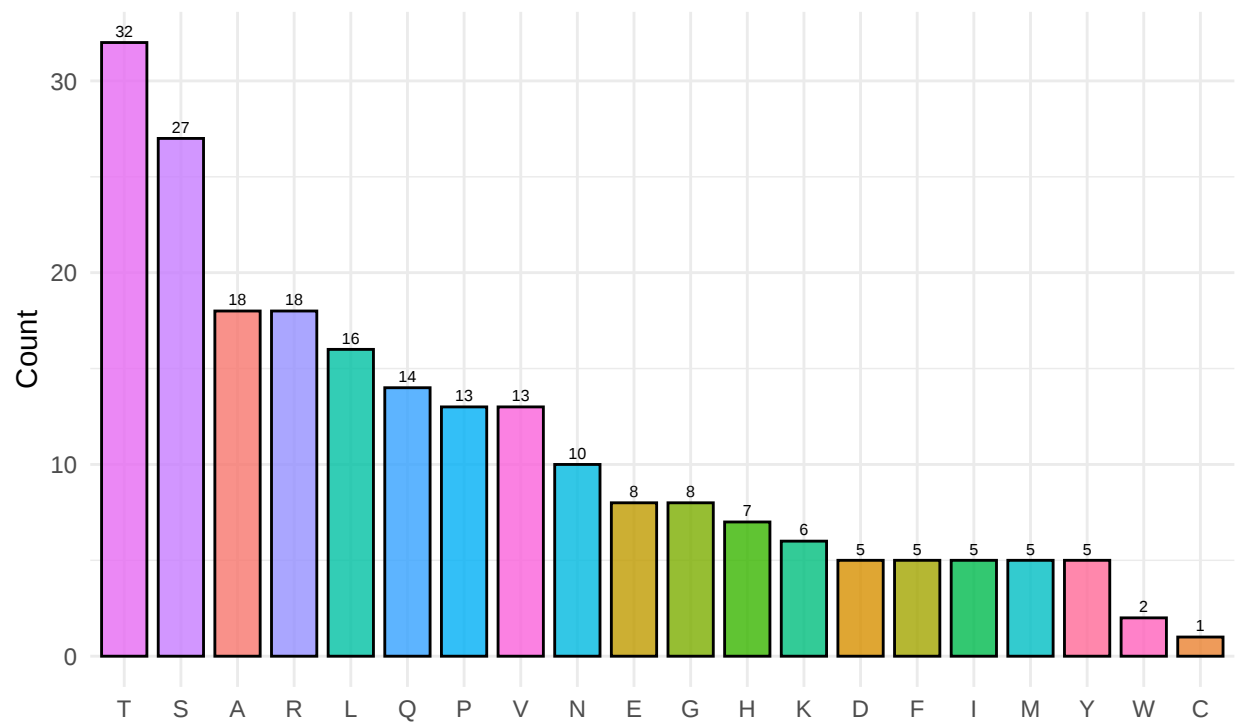
Amoniacid count Position 15



Position 16:

```
data1 %>%
  dplyr::count(x16) %>%
  ggplot(aes(x = reorder(x16, desc(n)), y = n, fill = x16)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 16",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

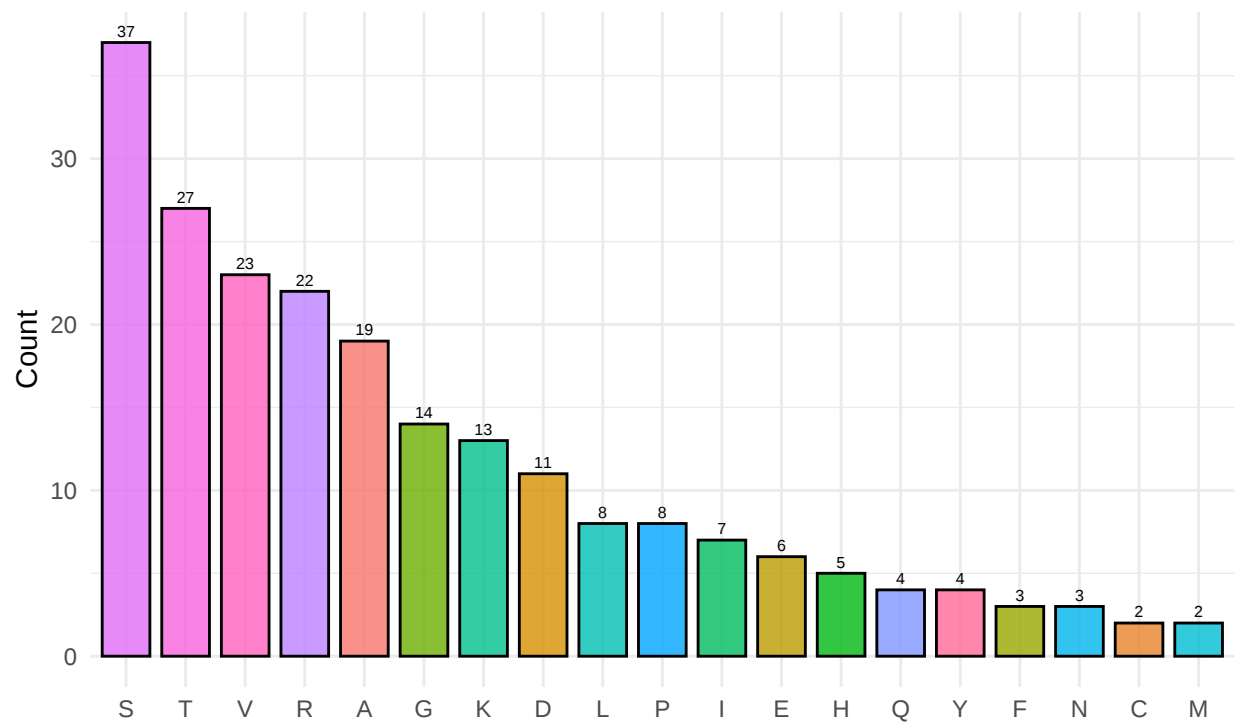
Amoniacid count Position 16



Position 17:

```
data1 %>%
  dplyr::count(x17) %>%
  ggplot(aes(x = reorder(x17, desc(n)), y = n, fill = x17)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n, position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
    labs(title="Amoniacid count Position 17",
          subtitle = "",
          x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

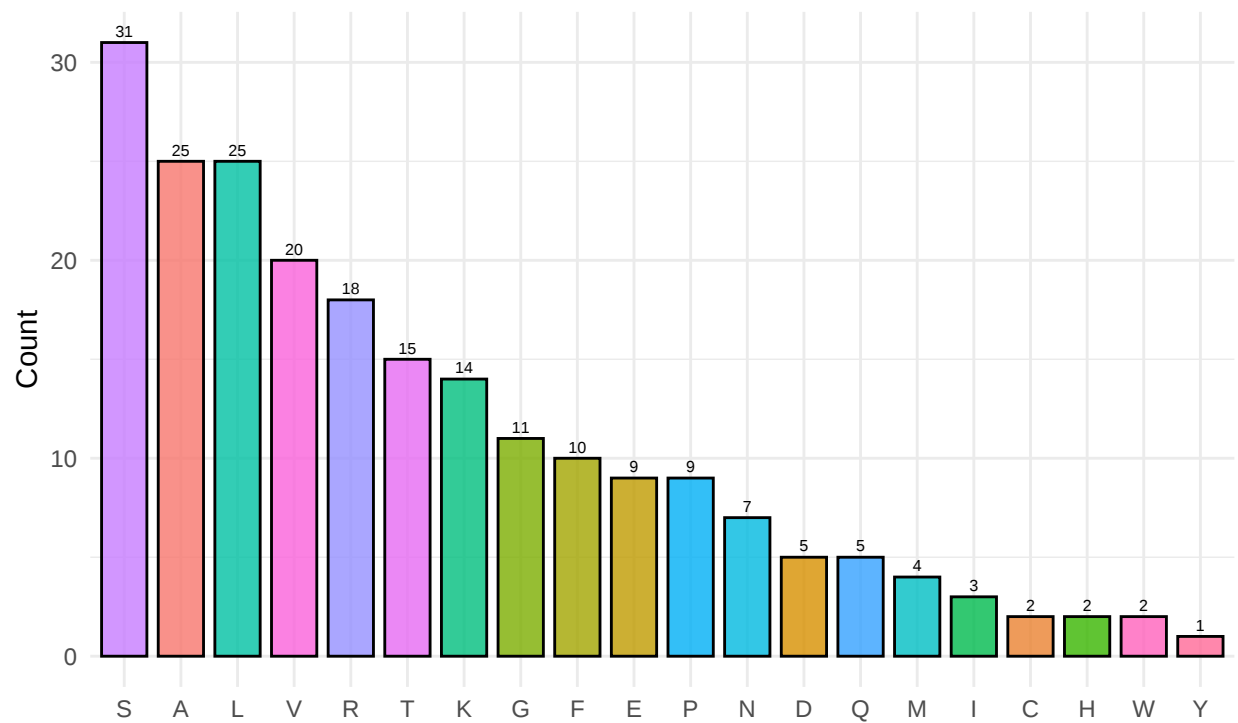
Amoniacid count Position 17



Position 18:

```
data1 %>%
  dplyr::count(x18) %>%
  ggplot(aes(x = reorder(x18, desc(n)), y = n, fill = x18)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 18",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

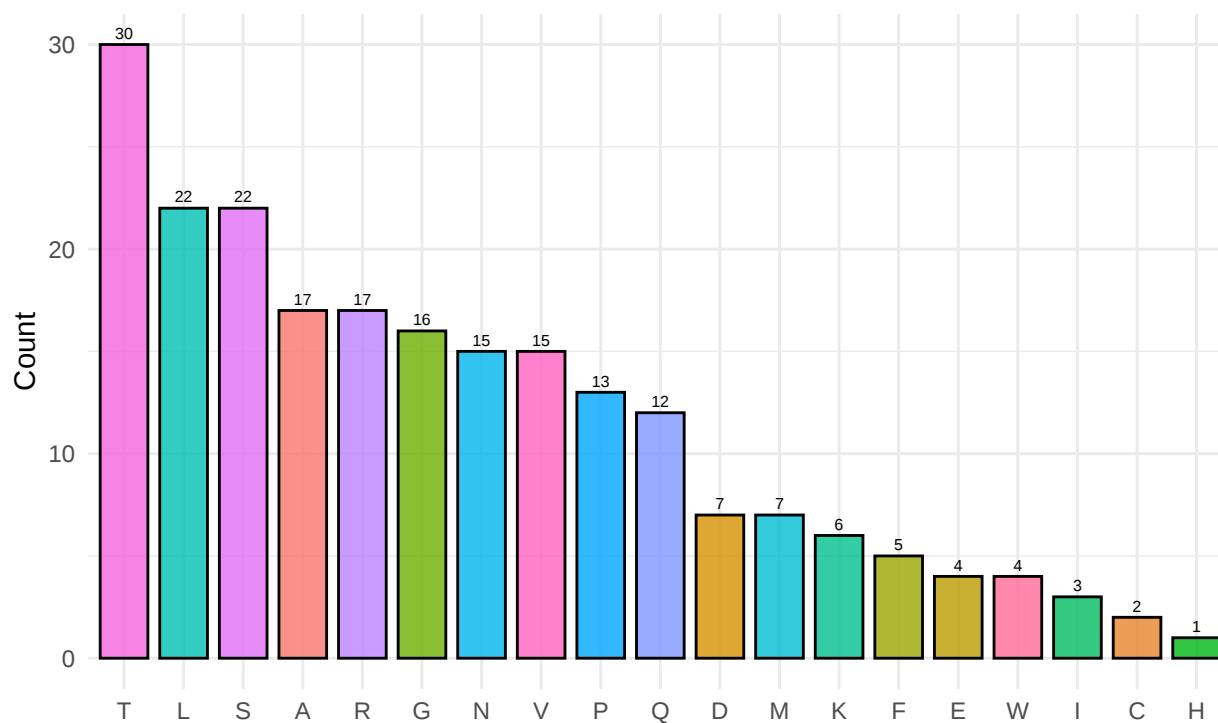
Amoniacid count Position 18



Position 19:

```
data1 %>%
  dplyr::count(x19) %>%
  ggplot(aes(x = reorder(x19, desc(n)), y = n, fill = x19)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 19",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

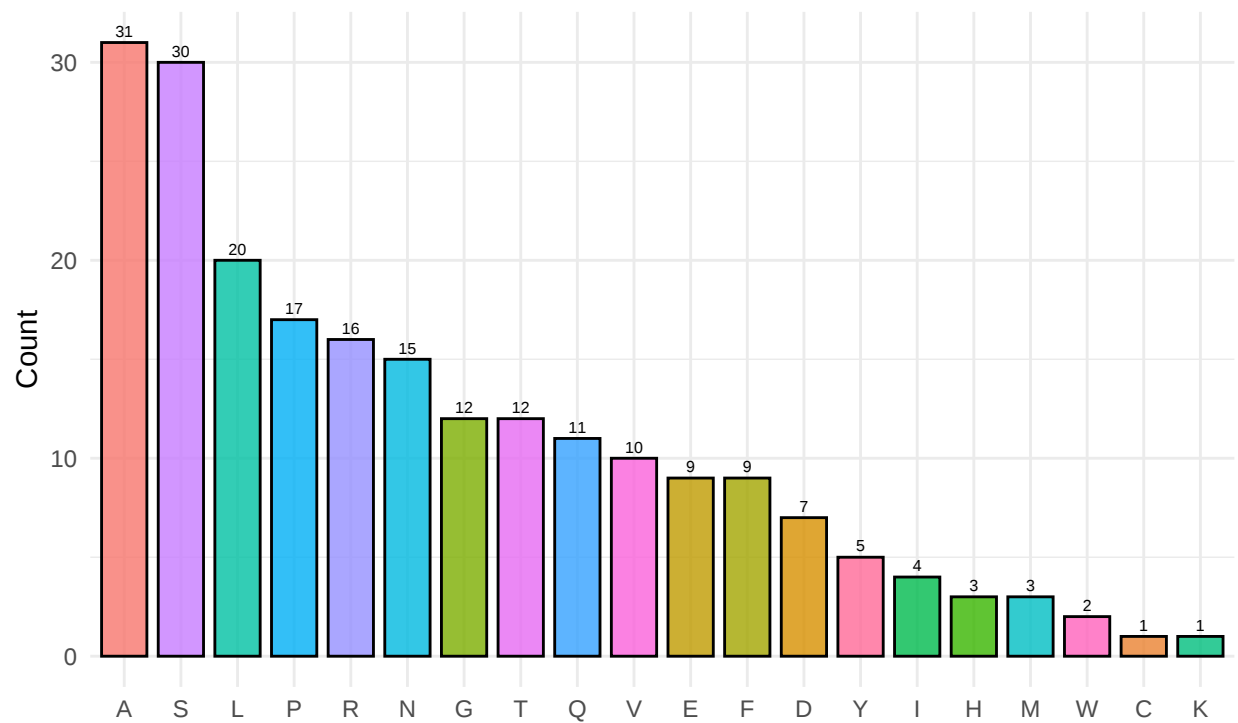
Amoniacid count Position 19



Position 20:

```
data1 %>%
  dplyr::count(x20) %>%
  ggplot(aes(x = reorder(x20, desc(n)), y = n, fill = x20)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 20",
        subtitle = "",
        x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

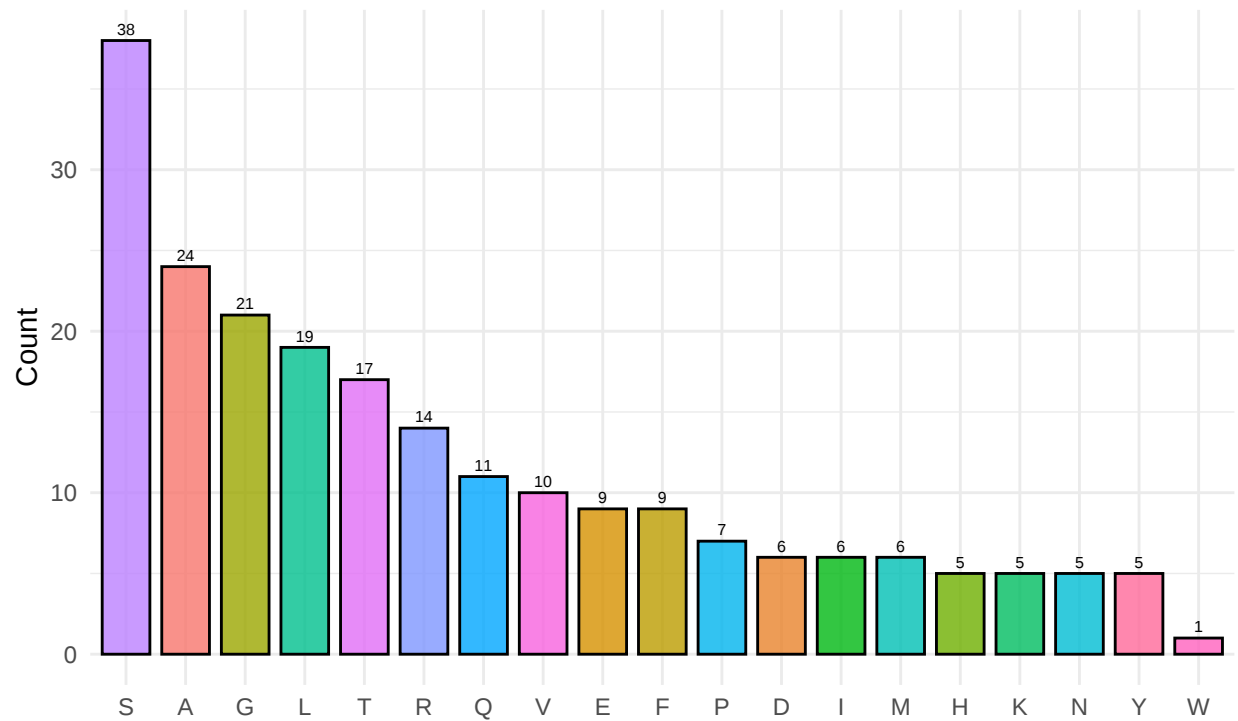
Amoniacid count Position 20



Position 21:

```
data1 %>%
  dplyr::count(x21) %>%
  ggplot(aes(x = reorder(x21, desc(n)), y = n, fill = x21)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 21",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

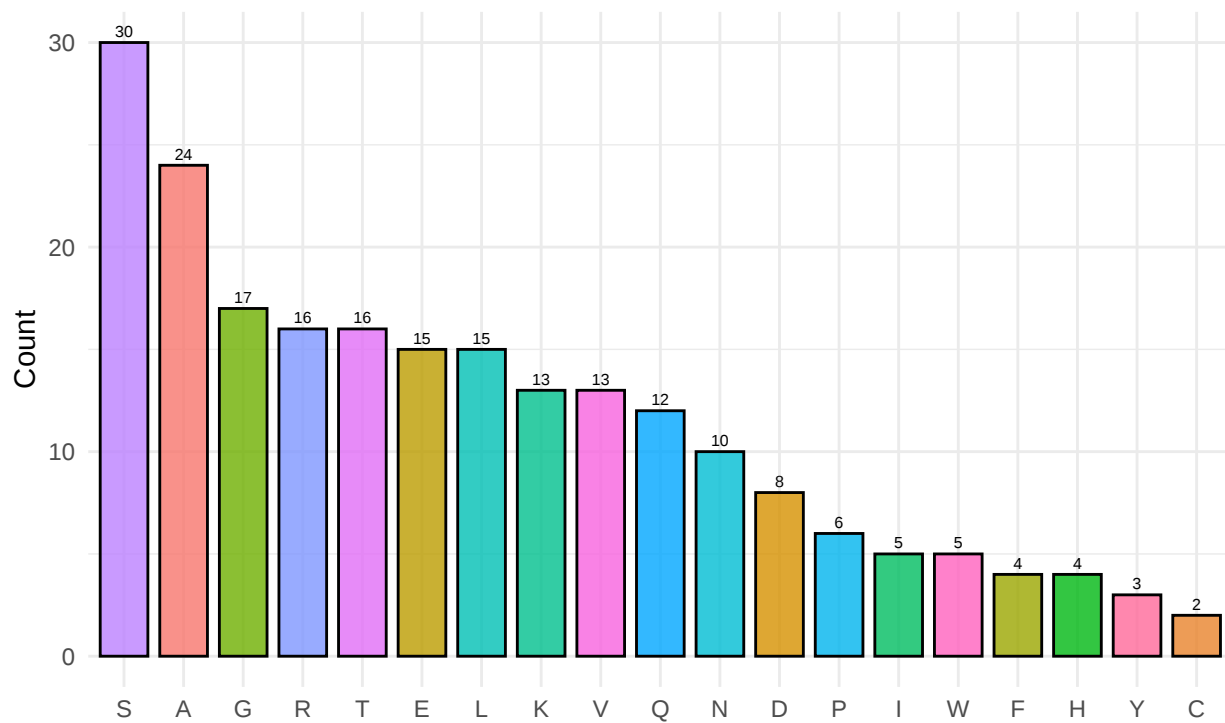
Amoniacid count Position 21



Position 22:

```
data1 %>%
  dplyr::count(x22) %>%
  ggplot(aes(x = reorder(x22, desc(n)), y = n, fill = x22)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 22",
        subtitle = "",
        x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

Amoniacid count Position 22



Position 23:

```
data1 %>%
  dplyr::count(x23) %>%
  ggplot(aes(x = reorder(x23, desc(n)), y = n, fill = x23)) +
  geom_bar(stat = "identity", position = "dodge", color = "black", alpha = 0.8, width = 0.8) +
  geom_text(aes(label=n), position=position_dodge(width=0.8), vjust= -0.5, size = 2) +
  labs(title="Amoniacid count Position 23",
       subtitle = "",
       x=" ", y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
```

Amoniacid count Position 23

