Hickey Lab!

Mariah Culpepper

```r
library(RColorBrewer)
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.0     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2
## ── Conflicts ──────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
to become errors
```

```r
library(dplyr)
library(ggplot2)
```

```r
data = read.csv("/Users/riahcul/Downloads/locofspatialdata.csv")
data[data == ""] <- NA
```

```
x_freqs <- data |>
  group_by(x_col_name) |>
  summarize(Frequency = n()) |>
  arrange(desc(Frequency))

y_freqs <- data |>
  group_by(y_col_name) |>
  summarize(Frequency = n()) |>
  arrange(desc(Frequency))

celltype_freqs <- data |>
  group_by(celltype_col_name) |>
  summarize(Frequency = n()) |>
  arrange(desc(Frequency))

uniqueregion_freqs <- data |>
  group_by(uniqueregion_col_name) |>
  summarize(Frequency = n()) |>
  arrange(desc(Frequency))

modality_freqs <- data |>
  group_by(modality) |>
  summarize(Frequency = n()) |>
  arrange(desc(Frequency))

disease_freqs <- data |>
  group_by(disease) |>
  summarize(Frequency = n()) |>
  arrange(desc(Frequency))
```

```
print(x_freqs)
```

```
## # A tibble: 25 × 2
##    x_col_name        Frequency
##    <chr>                 <int>
##  1 x                        15
##  2 X                        14
##  3 centroid-0                6
##  4 Location_Center_X         3
##  5 X_centroid                3
##  6 centroid_x                3
##  7 X:X                       2
##  8 X_cent                    2
##  9 Center_X                  1
## 10 NominalPostion_X          1
## # ℹ 15 more rows
```

```
print(y_freqs)
```

```
## # A tibble: 25 × 2
##    y_col_name        Frequency
##    <chr>                 <int>
##  1 y                        15
##  2 Y                        14
##  3 centroid-1                6
##  4 Location_Center_Y         3
##  5 Y_centroid                3
##  6 centroid_y                3
##  7 Y:Y                       2
##  8 Y_cent                    2
##  9 Center_Y                  1
## 10 NominalPosition_Y         1
## # ℹ 15 more rows
```

```
print(celltype_freqs)
```

```
## # A tibble: 33 × 2
##    celltype_col_name Frequency
##    <chr>                 <int>
##  1 <NA>                     10
##  2 cell_type                 6
##  3 phenotype                 5
##  4 cellType                  4
##  5 Cell Type                 3
##  6 Cluster                   3
##  7 ClusterName               3
##  8 celltype                  3
##  9 Annotation                2
## 10 CellType                  2
## # ℹ 23 more rows
```

```
print(uniqueregion_freqs)
```

```
## # A tibble: 31 × 2
##    uniqueregion_col_name Frequency
##    <chr>                     <int>
##  1 unique_region                11
##  2 fov                           8
##  3 region                        7
##  4 separated by File             4
##  5 Region                        3
##  6 roi                           3
##  7 <NA>                          3
##  8 frame                         2
##  9 imageID                       2
## 10 FileName                      1
## # ℹ 21 more rows
```

```
print(modality_freqs)
```

```
## # A tibble: 18 × 2
##    modality                       Frequency
##    <chr>                              <int>
##  1 CODEX                                 21
##  2 IMC                                   14
##  3 MIBI                                   8
##  4 CyCIF                                  4
##  5 CODEX,MIBI                             3
##  6 CosMx SMI                              3
##  7 3D Cell DIVE                           1
##  8 CODEX, multiplexed ISH                 1
##  9 CODEX,CyCIF,mIHC,MxIF,IMC,MIBI         1
## 10 IBEX                                   1
## 11 MACSima                                1
## 12 MALDI                                  1
## 13 MIF                                    1
## 14 MIF (using CODEX)                      1
## 15 Orion                                  1
## 16 Visium                                 1
## 17 mIHC                                   1
## 18 <NA>                                   1
```

```
print(disease_freqs)
```

```
## # A tibble: 9 × 2
##   disease             Frequency
##   <chr>                   <int>
## 1 Cancer                     39
## 2 Normal                     11
## 3 Normal,Cancer               4
## 4 Infection                   3
## 5 Inflammation                3
## 6 <NA>                        2
## 7 Normal,Inflammation         1
## 8 Osteoarthritis              1
## 9 Type 1 Diabetes             1
```
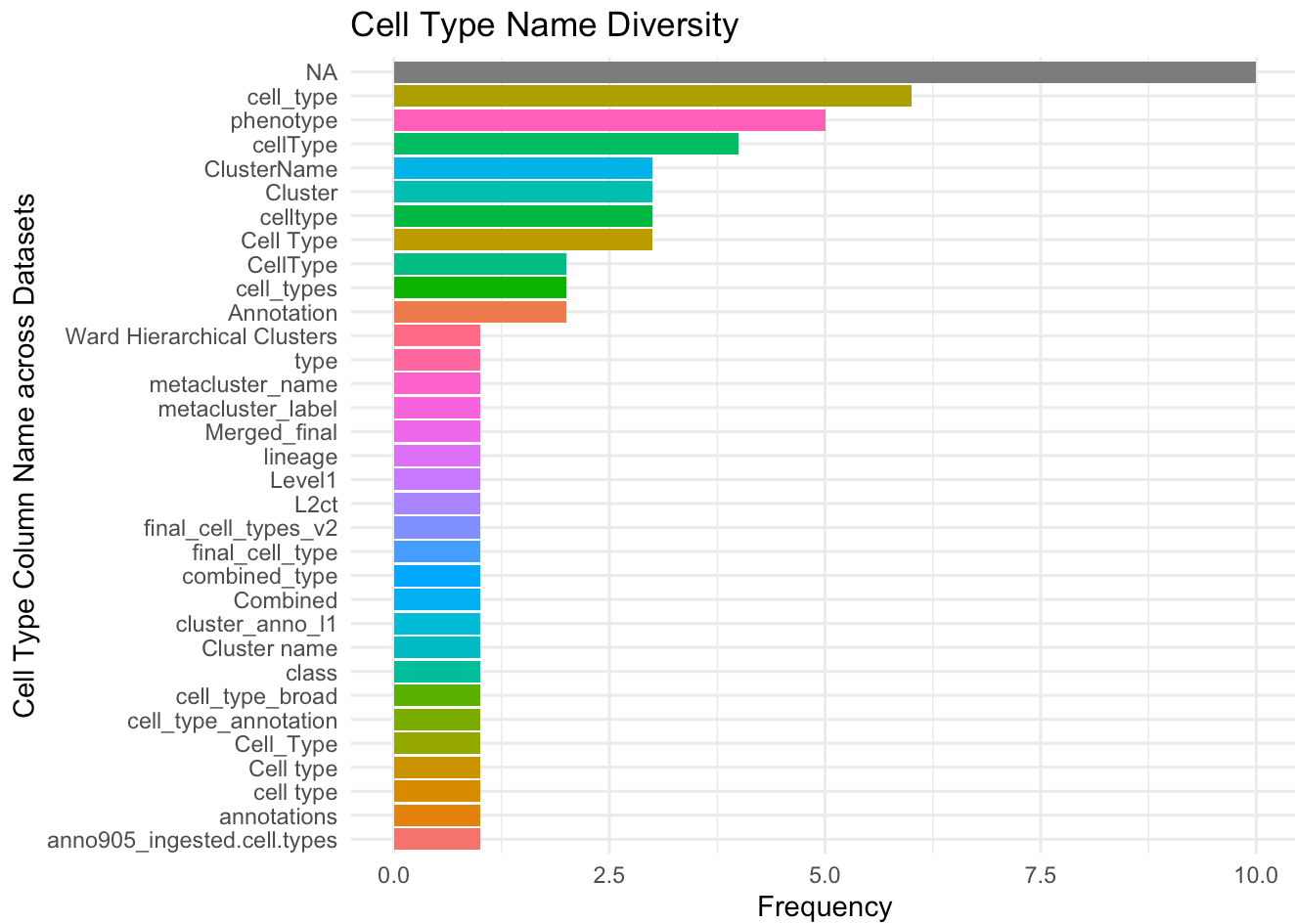
```
x_freqs |>
  ggplot(aes(x = reorder(x_col_name, Frequency), y=Frequency, fill = x_col_name)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(
    title = "X-Column Name Diversity",
    x = "X-Column Name across Datasets",
  ) +
  coord_flip()
```

X-Column Name Diversity
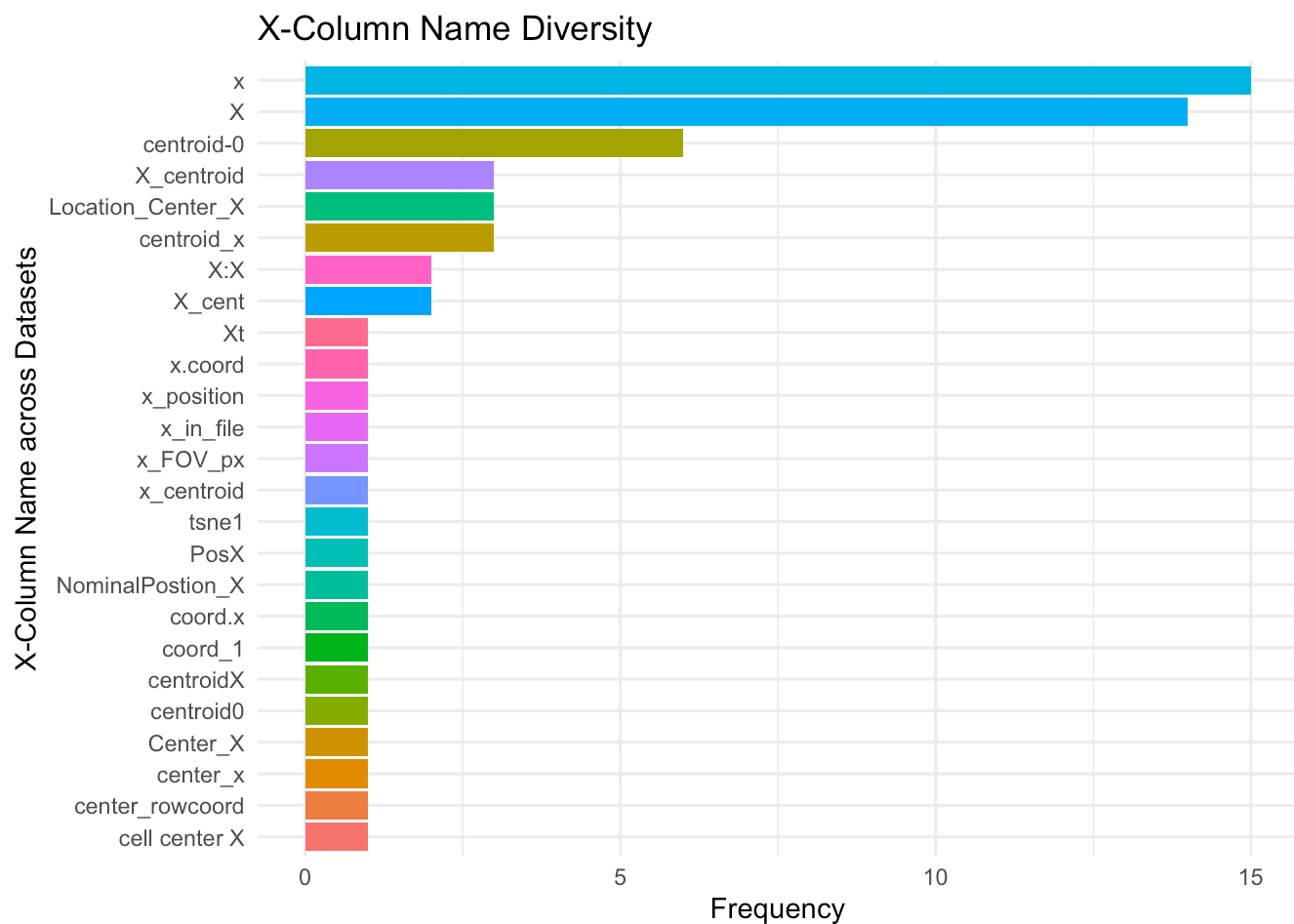

```
y_freqs |>
  ggplot(aes(x = reorder(y_col_name, Frequency), y=Frequency, fill = y_col_name)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(
    title = "Y—Column Name Diversity",
    x = "Y—Column Name across Datasets",
  ) +
  coord_flip()
```

Y-Column Name Diversity
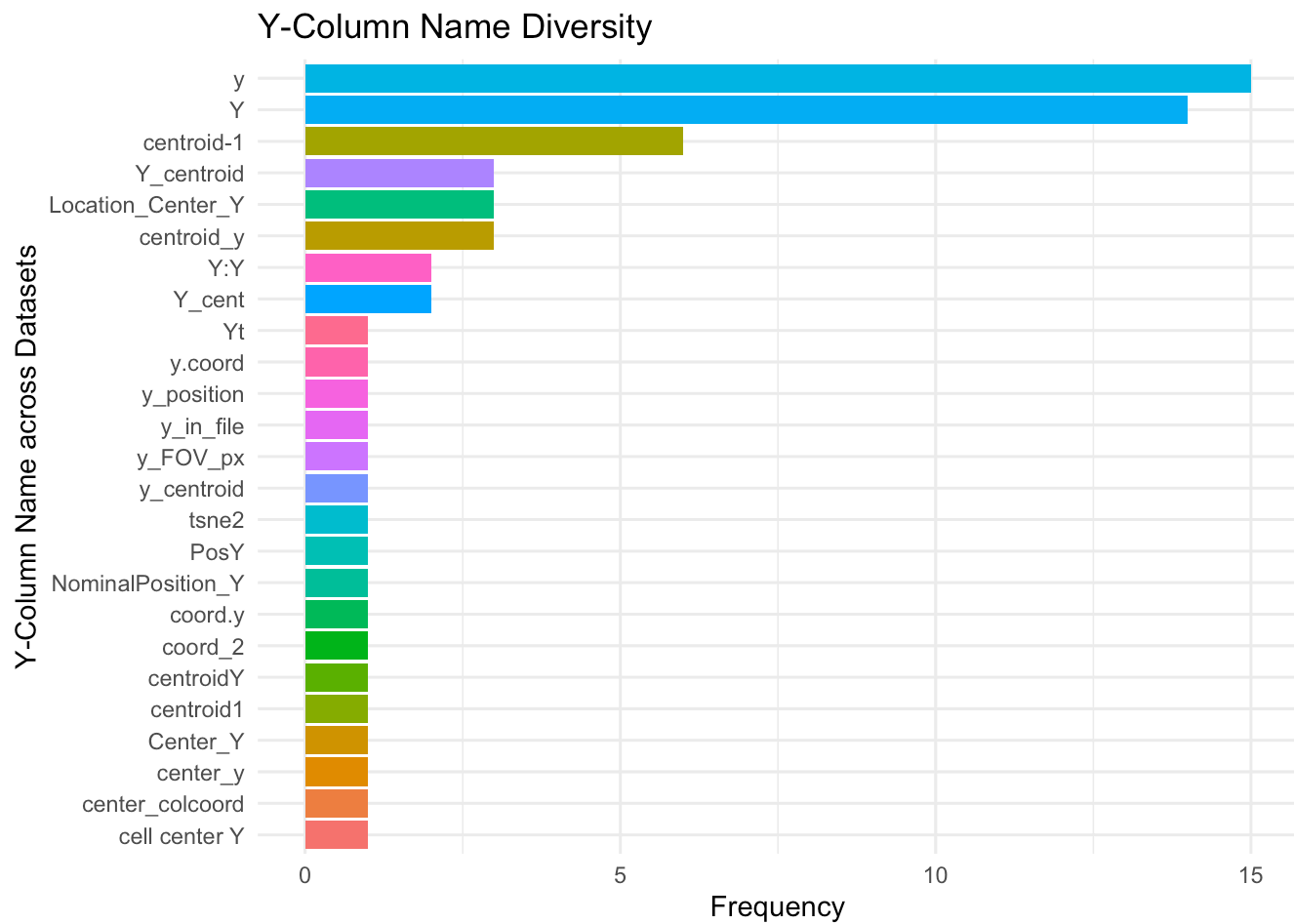

```
celltype_freqs |>
  ggplot(aes(x = reorder(celltype_col_name, Frequency), y=Frequency, fill = celltype_col
_name)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(
    title = "Cell Type Name Diversity",
    x = "Cell Type Column Name across Datasets",
  ) +
  coord_flip()
```

Cell Type Name Diversity


```
uniqueregion_freqs |>
  ggplot(aes(x = reorder(uniqueregion_col_name, Frequency), y=Frequency, fill = uniquere
gion_col_name)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(
    title = "Unique Region Name Diversity",
    x = "Unique Region Column Name across Datasets",
  ) +
  coord_flip()
```
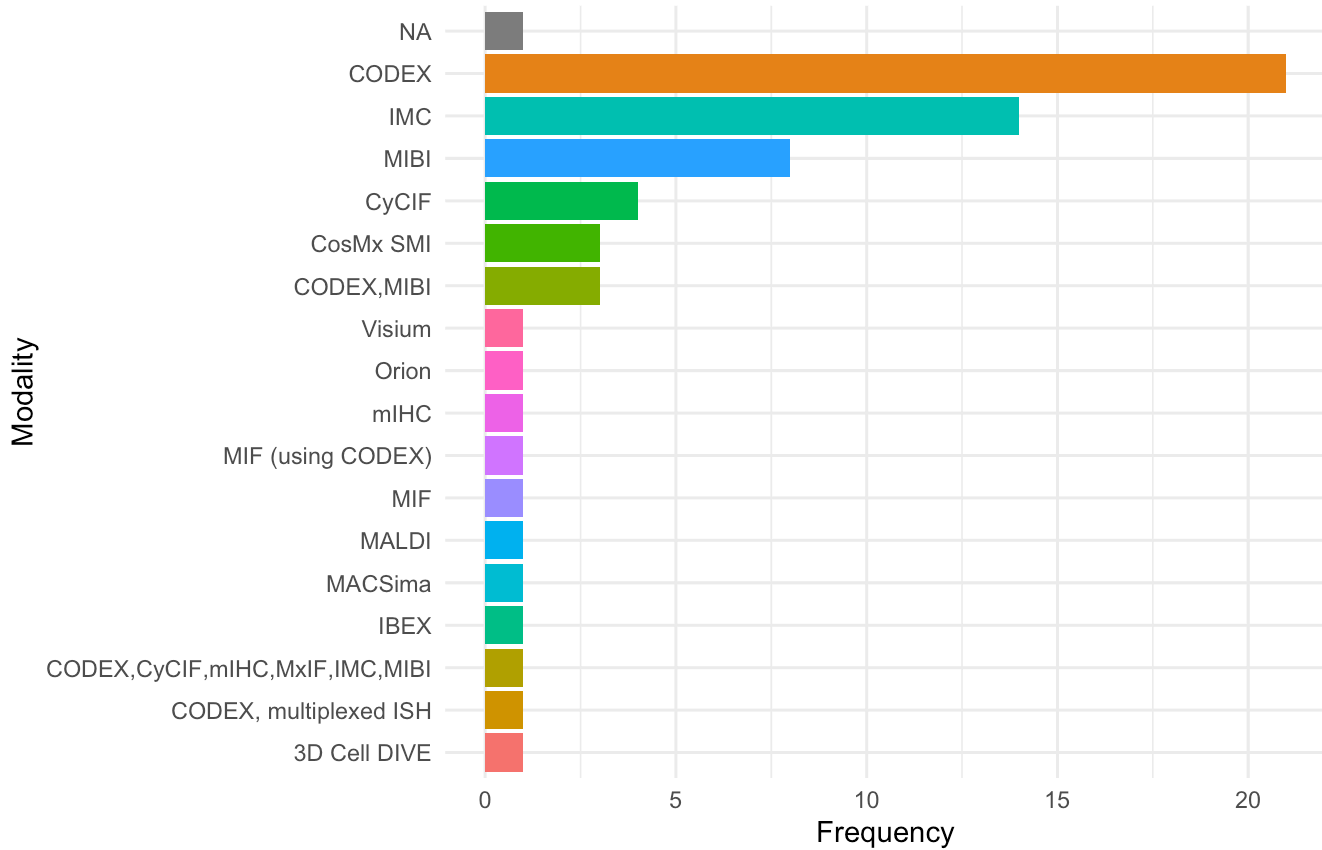
Unique Region Name Diversity
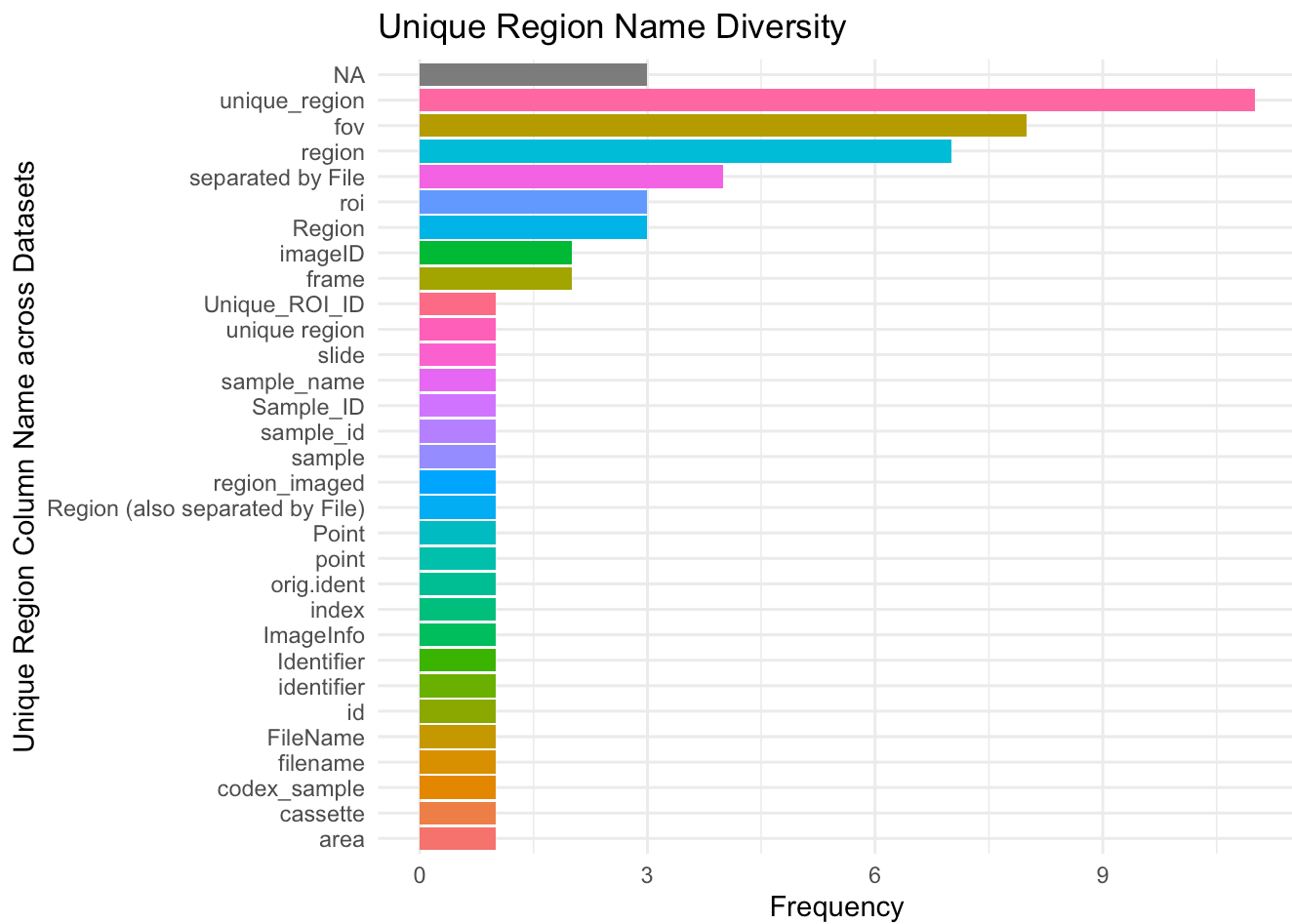

```
modality_freqs |>
  ggplot(aes(x = reorder(modality, Frequency), y=Frequency, fill = modality)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(
    title = "Variation in Modality",
    x = "Modality",
    subtitle = "Need to update"
  ) +
  coord_flip()
```

Variation in Modality
Need to update


```
disease_freqs |>
  ggplot(aes(x = reorder(disease, Frequency), y=Frequency, fill = disease)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(
    title = "Variation in Disease",
    x = "Diseases across Datasets",
  ) +
  coord_flip()
```

Variation in Disease