hackathon

```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)
library(gganimate)
```

```
## No renderer backend detected. gganimate will default to writing frames to separate fi
les
## Consider installing:
## - the `gifski` package for gif output
## - the `av` package for video output
## and restarting the R session
```

```r
# for the hackathon, focused on dataset visualizations and cell atlas visualizations

metadata = read.csv("/Users/riahcul/Downloads/20241106_allmetadata.csv")
metadata[metadata == ""] <- NA

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

```r
colnames(metadata)
```

```
## [1] "dataset_name"      "region_imaged"      "age"
## [4] "sex"               "donor"              "tissue"
## [7] "disease"           "renamed_tissue"     "large_tissue_unit"
```

```r
colnames(spatialdata)
```

```
##  [1] "manuscript"            "title"                 "doi"
##  [4] "priority"              "folder_name"           "x_col_name"
##  [7] "y_col_name"            "celltype_col_name"     "uniqueregion_col_name"
## [10] "metadata_file_added"   "tissue_col_name_md"    "disease_col_name_md"
## [13] "donor_col_name_md"     "age_col_name"          "sex_col_name"
## [16] "has_tissue_data"       "donor_col_name_seg"    "donor_col_name_mdd"
## [19] "associated_md"         "email_reason"          "need_to_email"
## [22] "need_to_email_md"      "email_drafted"         "emailed"
## [25] "author_email"          "been_emailed"          "received_data"
## [28] "singlecell_data_loc"   "image_data_loc"        "priority_2"
## [31] "species"               "tissue"                "disease"
## [34] "modality"              "lab"                   "added_by"
## [37] "checked_by"            "person_downloaded"     "not_usable"
## [40] "rrocessing_notes"      "processed"
```

```
unique(metadata$tissue)
```

```
##   [1] "breast"              NA                    "bone marrow"
##   [4] "tonsil"              "spleen"              "lymph node"
##   [7] "head and neck"       "thymus"              "liver"
##  [10] "biliary system"      "stomach"             "colon"
##  [13] "pancreas"            "ovary"               "uterus"
##  [16] "placenta"            "muscle"              "kidney"
##  [19] "nerve"               "soft tissue"         "tendon"
##  [22] "bone"                "skin"                "prostate"
##  [25] "adrenal gland"       "testis"              "brain"
##  [28] "meninges"            "lung"                "parathyroid"
##  [31] "thyroid"             "pleura"              "musculoskeletal"
##  [34] "bladder"             "vertebral body t7"   "decidua"
##  [37] "small intestine"     "esophagus"           "blood"
##  [40] "fallopiantube"       "ovarynormal"         "ovtp1"
##  [43] "ovtp10"              "ovtp100"             "ovtp101"
##  [46] "ovtp102"             "ovtp103"             "ovtp104"
##  [49] "ovtp105"             "ovtp106"             "ovtp107"
##  [52] "ovtp108"             "ovtp109"             "ovtp11"
##  [55] "ovtp110"             "ovtp111"             "ovtp112"
##  [58] "ovtp113"             "ovtp116"             "ovtp118"
##  [61] "ovtp119"             "ovtp120"             "ovtp121"
##  [64] "ovtp122"             "ovtp123"             "ovtp124"
##  [67] "ovtp125"             "ovtp127"             "ovtp13"
##  [70] "ovtp15"              "ovtp16"              "ovtp17"
##  [73] "ovtp18"              "ovtp19"              "ovtp2"
##  [76] "ovtp20"              "ovtp21"              "ovtp22"
##  [79] "ovtp24"              "ovtp26"              "ovtp27"
##  [82] "ovtp28"              "ovtp29"              "ovtp3"
##  [85] "ovtp30"              "ovtp31"              "ovtp32"
##  [88] "ovtp33"              "ovtp34"              "ovtp35"
##  [91] "ovtp36"              "ovtp37"              "ovtp38"
##  [94] "ovtp39"              "ovtp4"               "ovtp40"
##  [97] "ovtp41"              "ovtp42"              "ovtp43"
## [100] "ovtp44"              "ovtp45"              "ovtp46"
## [103] "ovtp47"              "ovtp49"              "ovtp5"
## [106] "ovtp50"              "ovtp51"              "ovtp53"
## [109] "ovtp54"              "ovtp55"              "ovtp56"
## [112] "ovtp58"              "ovtp59"              "ovtp6"
## [115] "ovtp60"              "ovtp61"              "ovtp63"
## [118] "ovtp64"              "ovtp65"              "ovtp66"
## [121] "ovtp67"              "ovtp68"              "ovtp69"
## [124] "ovtp7"               "ovtp70"              "ovtp71"
## [127] "ovtp72"              "ovtp73"              "ovtp74"
## [130] "ovtp75"              "ovtp76"              "ovtp77"
## [133] "ovtp78"              "ovtp79"              "ovtp8"
## [136] "ovtp80"              "ovtp81"              "ovtp82"
## [139] "ovtp83"              "ovtp84"              "ovtp85"
## [142] "ovtp87"              "ovtp88"              "ovtp89"
## [145] "ovtp9"               "ovtp90"              "ovtp91"
## [148] "ovtp92"              "ovtp93"              "ovtp95"
```

```
## [151] "ovtp96"              "ovtp97"              "ovtp98"
## [154] "ovtp99"
```

```
unique(metadata$disease)
```

```
##  [1] "cancer"              NA                    "normal"
##  [4] "non-malignant tumor" "ARDS"                "COVID19"
##  [7] "FLU"                 "Healthy"             "inflammation"
## [10] "tuberculosis"        "osteoarthritis"      "crc"
```

```
unique(metadata$large_tissue_unit)
```

```
##  [1] "Reproductive System"  NA                     "Musculoskeletal System"
##  [4] "Immune System"        "Digestive System"     "Urinary System"
##  [7] "Nervous System"       "Integumentary System" "Endocrine System"
## [10] "Respiratory System"   "Circulatory System"
```

```
# newmeta_update <- metadata |>
#   mutate(
#     newAge = round(as.numeric(age)))
#
# newmeta_update <- newmeta_update |>
#   mutate(
#     disease = if_else(disease == "Healthy", "normal", disease)
#   )
#
# meta_select <- metadata[metadata$disease %in% c("cancer", "normal", "COVID19", "ARD
S")]
#
# meta_select |>
#   ggplot(
#     aes(x=age, y= ))
```

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

```
disease_freqs |>
  ggplot(aes(x = disease, y=Frequency)) +
  geom_point(size=3) +
  geom_segment(aes(x= disease, xend=disease, y=0, yend = Frequency)) +
  theme(legend.position = "none") +
  labs(
    title = "Variation in Disease",
    x = "Diseases across Datasets",
  ) + theme_minimal()
```

Variation in Disease