BDSI R Workshop - Applying Wrangling and Visualisation skills

Robert Cope & Terry Neeman

2022

library(tidyverse)  
library(janitor)  
library(gtsummary)

In this exercise we will explore a recent published dataset, from this paper:

Marcato, F., van den Brand, H., Jansen, C. A., Rutten, V. P., Kemp, B., Engel, B., … & van Reenen, K. (2021). Effects of pre-transport diet, transport duration and transport condition on immune cell subsets, haptoglobin, cortisol and bilirubin in young veal calves. PloS one, 16(2), e0246959.

The aim of this study was to investigate effects of pre-transport diets (rearing milk OR electrolytes), transport durations (6 OR 18 hours) and transport conditions (open OR conditioned truck) on immune cell subsets, haptoglobin, cortisol and bilirubin of young calves upon arrival at the veal farm.

In this document, we create a workflow. First we read in the data. Then we do some data wrangling to put the data in a format that’s easy to visualise. The third step in the workflow is data visualisation. Data visualisation using ggplot2 is an essential part of understanding relevant patterns in the data.

## Read and clean up the data

veal <- read\_csv('data/Marcato\_2021\_immunity.csv')

## Rows: 1024 Columns: 37  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (10): Timepoint, Uplo, Bafr, Diet, Type, Bilirubin\_1, Cortisol\_1, Haptog...  
## dbl (27): Batch, Animal, Eartag, Pen, Corridor, Duration, wbc, neutro1, lymp...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

glimpse(veal)

## Rows: 1,024  
## Columns: 37  
## $ Timepoint <chr> "CC", "CC", "CC", "CC", "CC", "CC", "CC", "CC", "…  
## $ Batch <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…  
## $ Animal <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15…  
## $ Eartag <dbl> 48454, 4796, 85085, 884, 59787, 73475, 73981, 454…  
## $ Uplo <chr> "up", "up", "low", "low", "up", "up", "low", "low…  
## $ Bafr <chr> "back", "back", "back", "back", "back", "back", "…  
## $ Pen <dbl> 4, 4, 6, 6, 7, 7, 2, 2, 10, 10, 13, 13, 14, 14, 1…  
## $ Corridor <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3…  
## $ Diet <chr> "E", "E", "M", "M", "E", "E", "M", "M", "E", "E",…  
## $ Duration <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6…  
## $ Type <chr> "U", "U", "C", "C", "C", "C", "U", "U", "C", "C",…  
## $ Bilirubin\_1 <chr> "12.95", "18.51", "26.11", "9.84", "19.05", "4.60…  
## $ Cortisol\_1 <chr> "7.58", "16.55", "17.49", "12.37", "3.55", "4.46"…  
## $ Haptoglobin <chr> "0.31", "0.28", "0.18", "0.55", "0.17", "0.15", "…  
## $ Titer\_IgG <chr> "4.7", "3.9", "4.4", "4.9", "5.3", "5.7", "5.4", …  
## $ Titer\_IgM <chr> "6.6", "4.0", "5.3", "3.8", "4.5", "5.7", "2.0", …  
## $ wbc <dbl> 6.79, 13.31, 11.68, 5.29, 6.87, 8.96, 9.78, 16.94…  
## $ neutro1 <dbl> 2.77, 6.02, 4.17, 1.60, 2.22, 4.94, 3.00, 8.65, 3…  
## $ lympho1 <dbl> 2.69, 4.75, 3.80, 2.33, 3.58, 1.79, 4.65, 5.55, 4…  
## $ mono1 <dbl> 1.15, 1.87, 2.20, 1.04, 0.90, 1.10, 1.59, 2.07, 1…  
## $ eo1 <dbl> 0.07, 0.16, 0.46, 0.14, 0.08, 0.73, 0.41, 0.51, 0…  
## $ baso1 <dbl> 0.11, 0.52, 1.06, 0.18, 0.10, 0.40, 0.14, 0.17, 0…  
## $ neutro2 <dbl> 40.80, 45.30, 35.60, 30.20, 32.30, 55.15, 30.70, …  
## $ lympho2 <dbl> 39.65, 35.65, 32.55, 43.95, 52.10, 19.95, 47.55, …  
## $ mono2 <dbl> 16.95, 14.05, 18.80, 19.75, 13.05, 12.25, 16.20, …  
## $ eo2 <dbl> 1.05, 1.15, 3.95, 2.60, 1.15, 8.15, 4.15, 3.05, 0…  
## $ baso2 <dbl> 1.55, 3.85, 9.10, 3.50, 1.40, 4.50, 1.40, 1.00, 2…  
## $ `CD8+` <dbl> 7.05, 5.73, 10.70, 5.15, 8.08, 12.10, 15.50, 5.19…  
## $ `CD8+ perf+` <dbl> 0.67, 1.31, 1.41, 2.08, 0.78, 1.24, 0.00, 0.34, 0…  
## $ `NK+` <dbl> 4.52, 6.52, 13.30, 3.16, 14.40, 13.80, 8.30, 5.22…  
## $ `NK+ perf+` <dbl> 1.64, 2.86, 1.30, 1.43, 2.17, 0.94, 0.82, 1.39, 1…  
## $ `CD4+` <dbl> 17.90, 11.20, 0.10, 0.09, 8.47, 22.50, 14.90, 8.1…  
## $ `CD4+ perf+` <dbl> 0.02, 0.25, 4.17, 11.80, 0.19, 0.12, 0.00, 0.06, …  
## $ `gamma delta+` <dbl> 41.60, 48.90, 23.50, 55.90, 36.90, 18.90, 3.17, 6…  
## $ `gamma delta+ perf+` <dbl> 0.06, 0.33, 0.12, 0.23, 0.07, 0.19, 0.00, 0.12, 0…  
## $ `CD14+ large` <dbl> 35.20, 34.20, 44.70, 24.10, 19.10, 61.10, 3.93, 4…  
## $ `CD21+ large` <dbl> 8.04, 12.50, 12.80, 6.85, 16.30, 4.30, 29.10, 3.4…

Many of the later columns here have spaces or punctuation that is awkward to work with. A convenient way of dealing with this is to use the clean\_names() function from the janitor package.

veal <- clean\_names(veal)  
glimpse(veal)

## Rows: 1,024  
## Columns: 37  
## $ timepoint <chr> "CC", "CC", "CC", "CC", "CC", "CC", "CC", "CC", "CC",…  
## $ batch <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ animal <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16…  
## $ eartag <dbl> 48454, 4796, 85085, 884, 59787, 73475, 73981, 45400, …  
## $ uplo <chr> "up", "up", "low", "low", "up", "up", "low", "low", "…  
## $ bafr <chr> "back", "back", "back", "back", "back", "back", "back…  
## $ pen <dbl> 4, 4, 6, 6, 7, 7, 2, 2, 10, 10, 13, 13, 14, 14, 15, 1…  
## $ corridor <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3,…  
## $ diet <chr> "E", "E", "M", "M", "E", "E", "M", "M", "E", "E", "M"…  
## $ duration <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,…  
## $ type <chr> "U", "U", "C", "C", "C", "C", "U", "U", "C", "C", "U"…  
## $ bilirubin\_1 <chr> "12.95", "18.51", "26.11", "9.84", "19.05", "4.60", "…  
## $ cortisol\_1 <chr> "7.58", "16.55", "17.49", "12.37", "3.55", "4.46", "2…  
## $ haptoglobin <chr> "0.31", "0.28", "0.18", "0.55", "0.17", "0.15", "0.20…  
## $ titer\_ig\_g <chr> "4.7", "3.9", "4.4", "4.9", "5.3", "5.7", "5.4", "5.2…  
## $ titer\_ig\_m <chr> "6.6", "4.0", "5.3", "3.8", "4.5", "5.7", "2.0", "6.9…  
## $ wbc <dbl> 6.79, 13.31, 11.68, 5.29, 6.87, 8.96, 9.78, 16.94, 10…  
## $ neutro1 <dbl> 2.77, 6.02, 4.17, 1.60, 2.22, 4.94, 3.00, 8.65, 3.81,…  
## $ lympho1 <dbl> 2.69, 4.75, 3.80, 2.33, 3.58, 1.79, 4.65, 5.55, 4.87,…  
## $ mono1 <dbl> 1.15, 1.87, 2.20, 1.04, 0.90, 1.10, 1.59, 2.07, 1.48,…  
## $ eo1 <dbl> 0.07, 0.16, 0.46, 0.14, 0.08, 0.73, 0.41, 0.51, 0.06,…  
## $ baso1 <dbl> 0.11, 0.52, 1.06, 0.18, 0.10, 0.40, 0.14, 0.17, 0.23,…  
## $ neutro2 <dbl> 40.80, 45.30, 35.60, 30.20, 32.30, 55.15, 30.70, 51.0…  
## $ lympho2 <dbl> 39.65, 35.65, 32.55, 43.95, 52.10, 19.95, 47.55, 32.7…  
## $ mono2 <dbl> 16.95, 14.05, 18.80, 19.75, 13.05, 12.25, 16.20, 12.2…  
## $ eo2 <dbl> 1.05, 1.15, 3.95, 2.60, 1.15, 8.15, 4.15, 3.05, 0.55,…  
## $ baso2 <dbl> 1.55, 3.85, 9.10, 3.50, 1.40, 4.50, 1.40, 1.00, 2.20,…  
## $ cd8 <dbl> 7.05, 5.73, 10.70, 5.15, 8.08, 12.10, 15.50, 5.19, 14…  
## $ cd8\_perf <dbl> 0.67, 1.31, 1.41, 2.08, 0.78, 1.24, 0.00, 0.34, 0.86,…  
## $ nk <dbl> 4.52, 6.52, 13.30, 3.16, 14.40, 13.80, 8.30, 5.22, 8.…  
## $ nk\_perf <dbl> 1.64, 2.86, 1.30, 1.43, 2.17, 0.94, 0.82, 1.39, 1.66,…  
## $ cd4 <dbl> 17.90, 11.20, 0.10, 0.09, 8.47, 22.50, 14.90, 8.19, 2…  
## $ cd4\_perf <dbl> 0.02, 0.25, 4.17, 11.80, 0.19, 0.12, 0.00, 0.06, 0.49…  
## $ gamma\_delta <dbl> 41.60, 48.90, 23.50, 55.90, 36.90, 18.90, 3.17, 63.50…  
## $ gamma\_delta\_perf <dbl> 0.06, 0.33, 0.12, 0.23, 0.07, 0.19, 0.00, 0.12, 0.00,…  
## $ cd14\_large <dbl> 35.20, 34.20, 44.70, 24.10, 19.10, 61.10, 3.93, 43.70…  
## $ cd21\_large <dbl> 8.04, 12.50, 12.80, 6.85, 16.30, 4.30, 29.10, 3.40, 1…

When we look at this output, we notice that some of the columns that we expect to be numbers, are not. If you look at the csv you notice that they have ‘.’ instead of empty cells where data are missing. Converting these to numeric will work (and turn those dots into NAs), but it will show a warning.

For example:

veal$bilirubin\_1 <- as.numeric(veal$bilirubin\_1)

## Warning: NAs introduced by coercion

You can modify multiple columns at once using the function combination mutate(across()). Here is a helpful reference: <https://dplyr.tidyverse.org/articles/colwise.html>

veal <- veal %>%  
 mutate(across(c(cortisol\_1:titer\_ig\_m), as.numeric))

## Warning in mask$eval\_all\_mutate(quo): NAs introduced by coercion  
  
## Warning in mask$eval\_all\_mutate(quo): NAs introduced by coercion  
  
## Warning in mask$eval\_all\_mutate(quo): NAs introduced by coercion  
  
## Warning in mask$eval\_all\_mutate(quo): NAs introduced by coercion

## Initial checks of experimental setup

(reminder what our data now looks like)

glimpse(veal)

## Rows: 1,024  
## Columns: 37  
## $ timepoint <chr> "CC", "CC", "CC", "CC", "CC", "CC", "CC", "CC", "CC",…  
## $ batch <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ animal <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16…  
## $ eartag <dbl> 48454, 4796, 85085, 884, 59787, 73475, 73981, 45400, …  
## $ uplo <chr> "up", "up", "low", "low", "up", "up", "low", "low", "…  
## $ bafr <chr> "back", "back", "back", "back", "back", "back", "back…  
## $ pen <dbl> 4, 4, 6, 6, 7, 7, 2, 2, 10, 10, 13, 13, 14, 14, 15, 1…  
## $ corridor <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3,…  
## $ diet <chr> "E", "E", "M", "M", "E", "E", "M", "M", "E", "E", "M"…  
## $ duration <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,…  
## $ type <chr> "U", "U", "C", "C", "C", "C", "U", "U", "C", "C", "U"…  
## $ bilirubin\_1 <dbl> 12.95, 18.51, 26.11, 9.84, 19.05, 4.60, 8.35, 6.31, 7…  
## $ cortisol\_1 <dbl> 7.58, 16.55, 17.49, 12.37, 3.55, 4.46, 22.69, 5.84, 5…  
## $ haptoglobin <dbl> 0.31, 0.28, 0.18, 0.55, 0.17, 0.15, 0.20, 0.18, 0.27,…  
## $ titer\_ig\_g <dbl> 4.7, 3.9, 4.4, 4.9, 5.3, 5.7, 5.4, 5.2, 3.5, 3.9, 6.3…  
## $ titer\_ig\_m <dbl> 6.6, 4.0, 5.3, 3.8, 4.5, 5.7, 2.0, 6.9, 5.0, 4.8, 2.8…  
## $ wbc <dbl> 6.79, 13.31, 11.68, 5.29, 6.87, 8.96, 9.78, 16.94, 10…  
## $ neutro1 <dbl> 2.77, 6.02, 4.17, 1.60, 2.22, 4.94, 3.00, 8.65, 3.81,…  
## $ lympho1 <dbl> 2.69, 4.75, 3.80, 2.33, 3.58, 1.79, 4.65, 5.55, 4.87,…  
## $ mono1 <dbl> 1.15, 1.87, 2.20, 1.04, 0.90, 1.10, 1.59, 2.07, 1.48,…  
## $ eo1 <dbl> 0.07, 0.16, 0.46, 0.14, 0.08, 0.73, 0.41, 0.51, 0.06,…  
## $ baso1 <dbl> 0.11, 0.52, 1.06, 0.18, 0.10, 0.40, 0.14, 0.17, 0.23,…  
## $ neutro2 <dbl> 40.80, 45.30, 35.60, 30.20, 32.30, 55.15, 30.70, 51.0…  
## $ lympho2 <dbl> 39.65, 35.65, 32.55, 43.95, 52.10, 19.95, 47.55, 32.7…  
## $ mono2 <dbl> 16.95, 14.05, 18.80, 19.75, 13.05, 12.25, 16.20, 12.2…  
## $ eo2 <dbl> 1.05, 1.15, 3.95, 2.60, 1.15, 8.15, 4.15, 3.05, 0.55,…  
## $ baso2 <dbl> 1.55, 3.85, 9.10, 3.50, 1.40, 4.50, 1.40, 1.00, 2.20,…  
## $ cd8 <dbl> 7.05, 5.73, 10.70, 5.15, 8.08, 12.10, 15.50, 5.19, 14…  
## $ cd8\_perf <dbl> 0.67, 1.31, 1.41, 2.08, 0.78, 1.24, 0.00, 0.34, 0.86,…  
## $ nk <dbl> 4.52, 6.52, 13.30, 3.16, 14.40, 13.80, 8.30, 5.22, 8.…  
## $ nk\_perf <dbl> 1.64, 2.86, 1.30, 1.43, 2.17, 0.94, 0.82, 1.39, 1.66,…  
## $ cd4 <dbl> 17.90, 11.20, 0.10, 0.09, 8.47, 22.50, 14.90, 8.19, 2…  
## $ cd4\_perf <dbl> 0.02, 0.25, 4.17, 11.80, 0.19, 0.12, 0.00, 0.06, 0.49…  
## $ gamma\_delta <dbl> 41.60, 48.90, 23.50, 55.90, 36.90, 18.90, 3.17, 63.50…  
## $ gamma\_delta\_perf <dbl> 0.06, 0.33, 0.12, 0.23, 0.07, 0.19, 0.00, 0.12, 0.00,…  
## $ cd14\_large <dbl> 35.20, 34.20, 44.70, 24.10, 19.10, 61.10, 3.93, 43.70…  
## $ cd21\_large <dbl> 8.04, 12.50, 12.80, 6.85, 16.30, 4.30, 29.10, 3.40, 1…

At the Collection Centre, calves were randomly allocated by the manager of the collection centre to one of eight treatment groups (N = 23 calves per treatment group per batch). Within each batch, 8 calves per treatment group were randomly selected for blood sampling. If we just use the data from the Collection Centre, can we check this?

veal %>%  
 filter(timepoint == "CC")%>%  
 group\_by(batch, diet,duration,type) %>%  
 summarise(count = n()) %>%  
 pivot\_wider(names\_from = batch, values\_from = count)

## `summarise()` has grouped output by 'batch', 'diet', 'duration'. You can  
## override using the `.groups` argument.

## # A tibble: 8 × 5  
## # Groups: diet, duration [4]  
## diet duration type `1` `2`  
## <chr> <dbl> <chr> <int> <int>  
## 1 E 6 C 8 8  
## 2 E 6 U 8 8  
## 3 E 18 C 8 8  
## 4 E 18 U 8 8  
## 5 M 6 C 8 8  
## 6 M 6 U 8 8  
## 7 M 18 C 8 8  
## 8 M 18 U 8 8

It’s unclear from the data whether the animalID comes from the “eartag” or the “animal” column. In my mind, the “eartag” column sounds more like an ID name, but check out whether they are the same. From the code below, it appears that they don’t correspond at all! So for the future, we’ll use “eartag” as the animalID variable.

veal %>%  
 group\_by(eartag, animal) %>%  
 summarise(count = n()) %>%  
 filter(count > 0)

## `summarise()` has grouped output by 'eartag'. You can override using the  
## `.groups` argument.

## # A tibble: 401 × 3  
## # Groups: eartag [130]  
## eartag animal count  
## <dbl> <dbl> <int>  
## 1 93 17 2  
## 2 93 18 1  
## 3 93 41 5  
## 4 366 69 2  
## 5 366 70 1  
## 6 366 99 3  
## 7 366 126 2  
## 8 884 4 2  
## 9 884 6 3  
## 10 884 11 1  
## # … with 391 more rows

In our data set there are 8 rows per treatment per batch.(16 rows per treatment) Check that the column ‘eartag’ captures the data for each animal/timepoint.

data\_by\_eartag <- veal %>%  
 group\_by(eartag, timepoint) %>%  
 summarise(count = n()) %>%  
 pivot\_wider(names\_from = timepoint, values\_from = count)

## `summarise()` has grouped output by 'eartag'. You can override using the  
## `.groups` argument.

data\_by\_eartag

## # A tibble: 130 × 9  
## # Groups: eartag [130]  
## eartag CC T0 T24 T4 T48 `Week 1` `Week 3` `Week 5`  
## <dbl> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 93 1 1 1 1 1 1 1 1  
## 2 366 1 1 1 1 1 1 1 1  
## 3 884 1 1 1 1 1 1 1 1  
## 4 888 1 1 1 1 1 1 1 1  
## 5 895 1 1 1 1 1 1 1 1  
## 6 902 1 1 1 1 1 1 1 1  
## 7 907 1 1 1 1 1 1 1 1  
## 8 1564 1 1 1 1 1 1 1 1  
## 9 1624 1 1 1 1 1 1 1 1  
## 10 3568 1 1 1 1 1 1 1 1  
## # … with 120 more rows

Summarise immune markers for each treatment combination of diet & duration at T24:

veal %>%  
 filter(timepoint == "T24") %>%  
 mutate(treat = interaction(diet, duration)) %>%  
 select(treat, bilirubin\_1:titer\_ig\_m) %>%  
 tbl\_summary(by = treat)

## Table printed with {flextable}, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

| Characteristic | E.6, N = 321 | M.6, N = 321 | E.18, N = 321 | M.18, N = 321 |
| --- | --- | --- | --- | --- |
| bilirubin\_1 | 3.64 (2.86, 5.14) | 4.60 (2.65, 6.23) | 3.75 (2.78, 4.61) | 3.47 (2.65, 4.44) |
| Unknown | 0 | 0 | 1 | 0 |
| cortisol\_1 | 1.7 (0.7, 5.2) | 2.1 (0.7, 5.7) | 4.0 (1.1, 9.7) | 3.5 (0.7, 7.1) |
| Unknown | 2 | 7 | 3 | 5 |
| haptoglobin | 0.27 (0.23, 0.35) | 0.28 (0.21, 0.39) | 0.33 (0.24, 0.42) | 0.35 (0.26, 0.57) |
| Unknown | 0 | 0 | 1 | 0 |
| titer\_ig\_g | 4.30 (3.48, 5.18) | 4.60 (3.20, 5.05) | 4.50 (3.60, 5.70) | 4.50 (3.08, 5.03) |
| Unknown | 0 | 0 | 1 | 0 |
| titer\_ig\_m | 4.70 (3.88, 5.95) | 4.70 (3.85, 5.93) | 4.95 (3.70, 5.53) | 5.15 (4.70, 5.82) |
| 1Median (IQR) | | | | |

## Exploratory plots

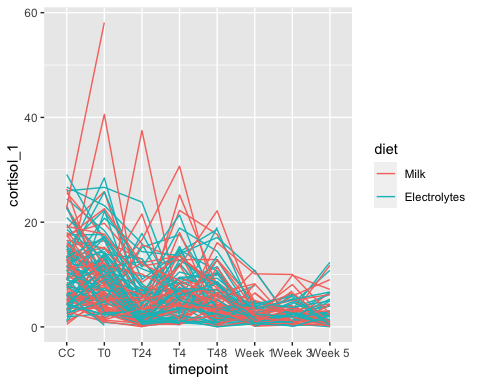
We approach data exploration from a research question perspective. The main questions of interest were in understanding the effect of calf handling/diet on immune cell subsets, haptoglobin, cortisol and bilirubin of young calves upon arrival at the veal farm. In this workflow, we focus on cortisol and CD8 cells. Similar exploratory plots can be done for the other markers.

What is the effect of diet on cortisol levels? At this point, we may want to indicate that M = milk and E = electrolytes. We can also make M the reference level.

veal <- mutate(veal, diet = factor(diet,   
 levels = c("M","E"),  
 labels = c("Milk","Electrolytes")))

ggplot(veal, aes(x=timepoint, y = cortisol\_1,colour = diet))+  
 geom\_line(aes(group = eartag))

## Warning: Removed 63 row(s) containing missing values (geom\_path).

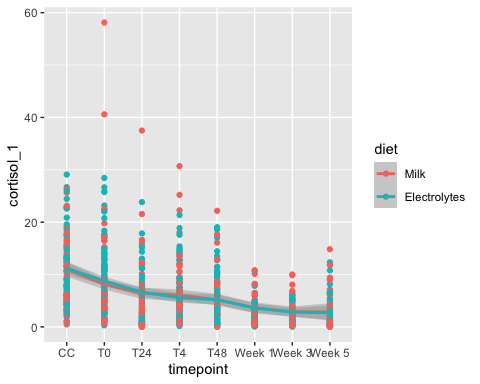


ggplot(veal, aes(x=timepoint, y = cortisol\_1, group = diet,colour = diet))+  
 geom\_point() +  
 geom\_smooth()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 152 rows containing non-finite values (stat\_smooth).

## Warning: Removed 152 rows containing missing values (geom\_point).



What do you notice is wrong with this plot?

The x axis is in the wrong order! you can fix this using factor() and specifying the levels.

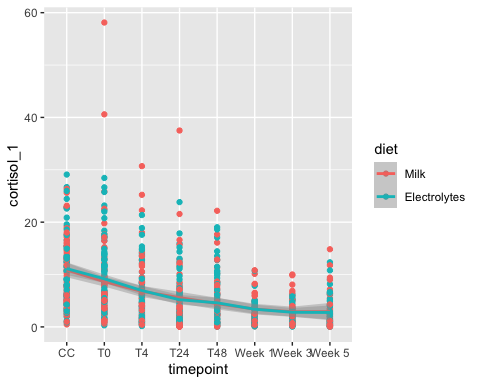
Fix the factor levels and remake the plot (just paste the plot code again).

veal$timepoint <- factor(veal$timepoint, levels = c('CC',  
 'T0',  
 'T4',  
 'T24',  
 'T48',  
 'Week 1',  
 'Week 3',  
 'Week 5'))  
  
ggplot(veal, aes(x=timepoint, y = cortisol\_1, group = diet, colour = diet))+  
 geom\_point() +   
 geom\_smooth()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 152 rows containing non-finite values (stat\_smooth).

## Warning: Removed 152 rows containing missing values (geom\_point).



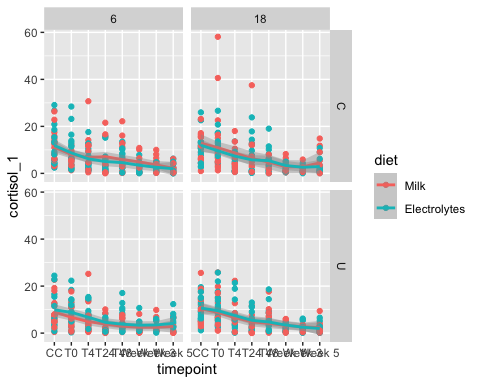
There may be a diet effect at the early time points, but there is a lot of variation in the data. Split these data into the 4 subplots: one for each transport and transport time factor.

ggplot(veal, aes(x=timepoint, y = cortisol\_1, group=diet, colour = diet))+  
 geom\_point() +   
 geom\_smooth() +  
 facet\_grid(type~duration)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 152 rows containing non-finite values (stat\_smooth).

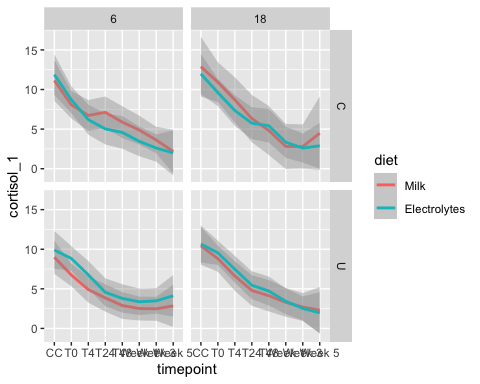
## Warning: Removed 152 rows containing missing values (geom\_point).

 We can also look at the smoothed curves without the data. This will change the y-axis limits. Does this perspective change your assessment?

ggplot(veal, aes(x=timepoint, y = cortisol\_1, group=diet, colour = diet))+  
 #geom\_point() +   
 geom\_smooth() +  
 facet\_grid(type~duration)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 152 rows containing non-finite values (stat\_smooth).



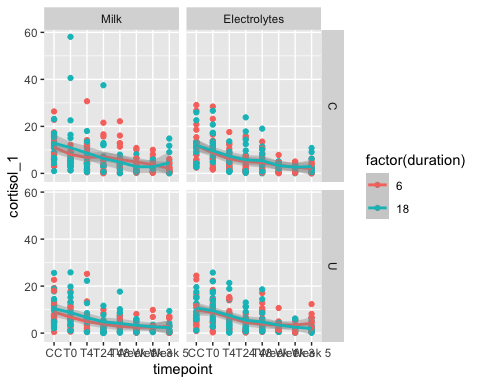
Now look at the duration of transport effect for each diet:transport type combination.

ggplot(veal, aes(x=timepoint, y = cortisol\_1, group=factor(duration), colour = factor(duration)))+  
 geom\_point() +   
 geom\_smooth() +  
 facet\_grid(type~diet)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 152 rows containing non-finite values (stat\_smooth).

## Warning: Removed 152 rows containing missing values (geom\_point).



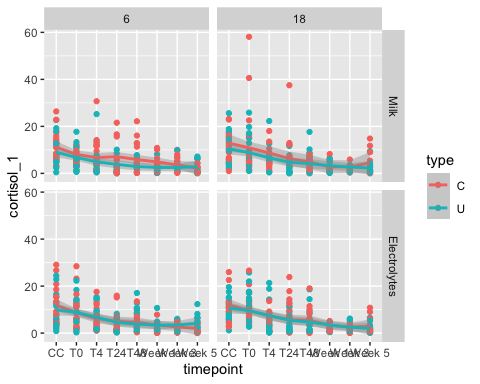
Finally, look at the effect of transport type on cortisol levels, for each diet and transport duration.

ggplot(veal, aes(x=timepoint, y = cortisol\_1, group=type, colour = type))+  
 geom\_point() +   
 geom\_smooth() +  
 facet\_grid(diet~duration)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 152 rows containing non-finite values (stat\_smooth).

## Warning: Removed 152 rows containing missing values (geom\_point).



What are the patterns you see in these data?

## FACS data

(reminder what our data now looks like)

str(veal)

## tibble [1,024 × 37] (S3: tbl\_df/tbl/data.frame)  
## $ timepoint : Factor w/ 8 levels "CC","T0","T4",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ batch : num [1:1024] 1 1 1 1 1 1 1 1 1 1 ...  
## $ animal : num [1:1024] 1 2 3 4 5 6 7 8 9 10 ...  
## $ eartag : num [1:1024] 48454 4796 85085 884 59787 ...  
## $ uplo : chr [1:1024] "up" "up" "low" "low" ...  
## $ bafr : chr [1:1024] "back" "back" "back" "back" ...  
## $ pen : num [1:1024] 4 4 6 6 7 7 2 2 10 10 ...  
## $ corridor : num [1:1024] 1 1 1 1 1 1 1 1 2 2 ...  
## $ diet : Factor w/ 2 levels "Milk","Electrolytes": 2 2 1 1 2 2 1 1 2 2 ...  
## $ duration : num [1:1024] 6 6 6 6 6 6 6 6 6 6 ...  
## $ type : chr [1:1024] "U" "U" "C" "C" ...  
## $ bilirubin\_1 : num [1:1024] 12.95 18.51 26.11 9.84 19.05 ...  
## $ cortisol\_1 : num [1:1024] 7.58 16.55 17.49 12.37 3.55 ...  
## $ haptoglobin : num [1:1024] 0.31 0.28 0.18 0.55 0.17 0.15 0.2 0.18 0.27 0.19 ...  
## $ titer\_ig\_g : num [1:1024] 4.7 3.9 4.4 4.9 5.3 5.7 5.4 5.2 3.5 3.9 ...  
## $ titer\_ig\_m : num [1:1024] 6.6 4 5.3 3.8 4.5 5.7 2 6.9 5 4.8 ...  
## $ wbc : num [1:1024] 6.79 13.31 11.68 5.29 6.87 ...  
## $ neutro1 : num [1:1024] 2.77 6.02 4.17 1.6 2.22 4.94 3 8.65 3.81 6.93 ...  
## $ lympho1 : num [1:1024] 2.69 4.75 3.8 2.33 3.58 1.79 4.65 5.55 4.87 3.77 ...  
## $ mono1 : num [1:1024] 1.15 1.87 2.2 1.04 0.9 1.1 1.59 2.07 1.48 1.09 ...  
## $ eo1 : num [1:1024] 0.07 0.16 0.46 0.14 0.08 0.73 0.41 0.51 0.06 0.15 ...  
## $ baso1 : num [1:1024] 0.11 0.52 1.06 0.18 0.1 0.4 0.14 0.17 0.23 0.19 ...  
## $ neutro2 : num [1:1024] 40.8 45.3 35.6 30.2 32.3 ...  
## $ lympho2 : num [1:1024] 39.6 35.6 32.5 44 52.1 ...  
## $ mono2 : num [1:1024] 16.9 14.1 18.8 19.8 13.1 ...  
## $ eo2 : num [1:1024] 1.05 1.15 3.95 2.6 1.15 8.15 4.15 3.05 0.55 1.2 ...  
## $ baso2 : num [1:1024] 1.55 3.85 9.1 3.5 1.4 4.5 1.4 1 2.2 1.5 ...  
## $ cd8 : num [1:1024] 7.05 5.73 10.7 5.15 8.08 12.1 15.5 5.19 14.4 11.5 ...  
## $ cd8\_perf : num [1:1024] 0.67 1.31 1.41 2.08 0.78 1.24 0 0.34 0.86 0.68 ...  
## $ nk : num [1:1024] 4.52 6.52 13.3 3.16 14.4 13.8 8.3 5.22 8.28 12.3 ...  
## $ nk\_perf : num [1:1024] 1.64 2.86 1.3 1.43 2.17 0.94 0.82 1.39 1.66 2.83 ...  
## $ cd4 : num [1:1024] 17.9 11.2 0.1 0.09 8.47 22.5 14.9 8.19 24.9 11.5 ...  
## $ cd4\_perf : num [1:1024] 0.02 0.25 4.17 11.8 0.19 0.12 0 0.06 0.49 0.12 ...  
## $ gamma\_delta : num [1:1024] 41.6 48.9 23.5 55.9 36.9 18.9 3.17 63.5 0.11 37 ...  
## $ gamma\_delta\_perf: num [1:1024] 0.06 0.33 0.12 0.23 0.07 0.19 0 0.12 0 0.09 ...  
## $ cd14\_large : num [1:1024] 35.2 34.2 44.7 24.1 19.1 61.1 3.93 43.7 28.3 25.1 ...  
## $ cd21\_large : num [1:1024] 8.04 12.5 12.8 6.85 16.3 4.3 29.1 3.4 11.5 6.92 ...

The FACS was only done at timepoints CC and T0, so we just want to work with those times.

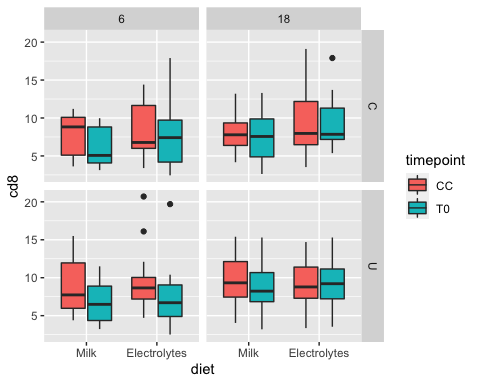
Filter the dataset to get just those timepoints, store this with the name veal\_start.

veal\_start <- veal %>% filter(timepoint %in% c('CC','T0'))

Plot a boxplot with these data, with x as timepoint, y as cd8, fill by diet and facet by duration and type.

ggplot(veal\_start, aes(x=diet, y=cd8,fill= timepoint))+  
 geom\_boxplot()+  
 facet\_grid(type~duration)

## Warning: Removed 1 rows containing non-finite values (stat\_boxplot).

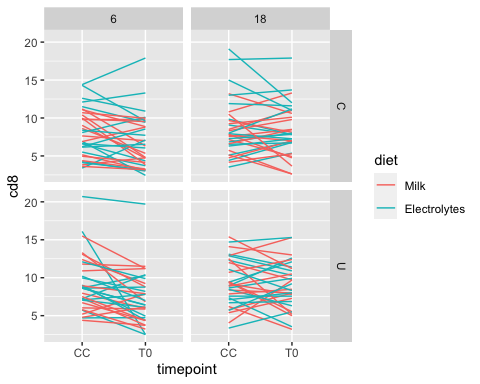


At the Collection Centre time, we expect the CD8 levels to have the same distribution for each treatment group since these are baseline measurements. If there were an effect of diet on CD8 levels, one would expect to see within-animal changes between CC and T0 to be different by diet.

Graph the within-animal changes over time.

ggplot(veal\_start, aes(x=timepoint, y=cd8,col=diet,group=eartag))+  
 geom\_line()+  
 facet\_grid(type~duration)

## Warning: Removed 1 row(s) containing missing values (geom\_path).



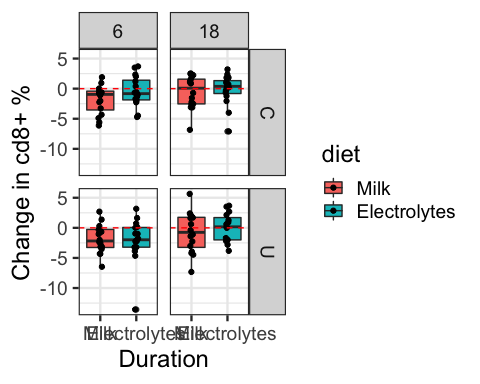
It does look like there is something happening here, but it is difficult to diagnose like this. Something that might be more informative is the change in cd8 between time CC and time T0.

We can do this, but it needs something a little complicated: a function called pivot\_wider(). We will look at how to do this in more detail in the next workshop, for now you should work from / edit this example:

veal\_diff <- veal\_start %>%  
 select(eartag,diet,duration,type,timepoint,cd8) %>%  
 pivot\_wider(names\_from = timepoint,  
 values\_from = cd8) %>%  
 mutate(change\_in\_cd8 = T0 - CC)  
  
 ggplot(veal\_diff, aes(x=diet,y=change\_in\_cd8, fill=diet))+  
 geom\_boxplot()+  
 geom\_point(position = position\_jitterdodge(jitter.width = 0.2))+  
 geom\_hline(yintercept = 0,col='red',lty='dashed')+  
 xlab('Duration')+ylab('Change in cd8+ %')+  
 theme\_bw(base\_size = 18)+  
 facet\_grid(type~duration)

## Warning: Removed 1 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 1 rows containing missing values (geom\_point).



## Exercise

Describe an experiment that your lab has done, including relevant background information.

Import your data and include some descriptive summaries.

Use ggplot to explore your research questions visually.