Stoebel Lab: Analysis of data from a single beta-gal experiment with multiple RpoS levels

This code imports, analyzes,and plots data from a single beta-galactosidase assay, as measured on our Bio-Tek plate reader. This analysis is for experiments examining beta-galactosidase activity as a function of RpoS level. If you are working with beta-galactosidase activity as a function of something else, this code will need to be modified. Talk to Dan.

Getting your data from the plate reader and into R.

First, export the data from Gen5.

Before importing into R, you need to do some preprocessing of the Excel file:

Leave Sheet 1 as it is. On Sheet2, copy and paste the OD420 data (i.e. the beta-gal assay itself). Include the whole matrix, even though we won't use most of it in R. Add seven more columns immediately to the right: experimenter, date, strain, RpoS, dilutionFactor, volumeAssayed, reactionTime. (note the capitalization and spaces)

Fill in **each** cell in these seven columns:

1. For **experimenter**, give your initials. Please always usethe same initials as you use for strain names.
2. For **date**, please give them in format YYYY-MM-DD, ie "2018-06-07". This date should correspond to the date noted in your lab notebook when this experiment was done.
3. For **strain**, please use your personal strain number (I.e LB32, ANS138, etc. . .). Wells used as a blank should be called "blank".
4. For **RpoS**, please give the % of wild-type in your condition, as determined by previous western blotting experiments. Wild-type = 100, knock-out = 0. Enter NA for wells used as a blank.
5. For **dilutionFactor**, if you diluted your cells 1:n before adding them to the reaction/plate, give the value of n. (E.g. if you diluted your cells 1:100 before adding a volume to be assayed, enter 100 here.) Enter NA for wells used as a blank.
6. For **volumeAssayed**, give the volume of (possibly diluted) cells in mL you added to the tube when you assayed. This is usually 0.1 or 0.5. Enter NA for wells used as a blank.
7. For **reactionTime**, give the length of time the reaction proceeded before adding stop solution in minutes. Enter NA for wells used as a blank.

On Sheet3, copy and paste the OD600 data (i.e. the cell densities). Add columns 1-5 above (experimenter, date, strain, RpoS, dilution factor) immediately to the right of the pasted data, with the exact same rules as above. The first four will be identical if you layed our your plate the same, but dilution factor will not. It is typically 10 for stationary phase cells.

Once you think you are done, *double-check everything.* Errors in data entry will mess everything else up.

Then, copy your data file to the folder/directory `myData`.

Getting ready to run the code

This code makes heavy use of tidyverse tools. FMI, I can't speak highly enough of the (free) book R for Data Science: http://r4ds.had.co.nz/

If this is the first time running this file, you may need to install the packages below. You can do this with the following command: `install.packages(c("readxl", "tidyverse")` . Once those packages are installed, you should be able to run the code.

```
library(readxl)
library(tidyverse)
theme_set(theme_classic())
source("betaGalAnalysisFunctions.R")
```

To start the analysis, read in the data.

```
fileName <- "myData/7.18.18 core.xlsx" #Enter your Excel file here. Remember that tab completion will h
OD420data <- read_excel(fileName, sheet = 2)
OD600data <- read_excel(fileName, sheet = 3)
```

Examine your data and make sure things look okay.

You need to run these two lines manually, as they will not be run by knitting

```
View(OD420data)
View(OD600data)
```

Calculate the measures of expression and cell density

We're going to start by constructing a new data frame that has just the elements we need for our analysis

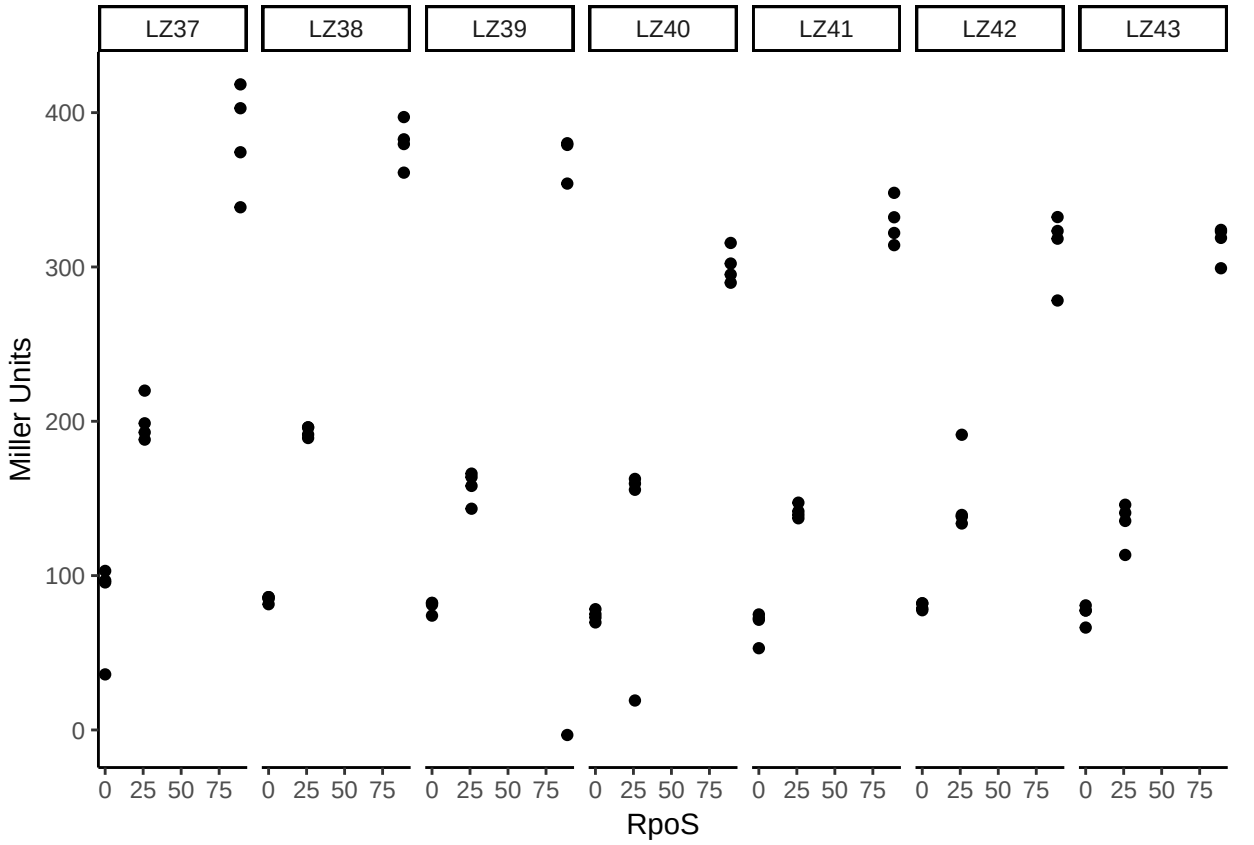
```
meanCellDensities <-  filter(OD600data, strain != "blank") %>%
                  mutate(trueDensity = `Blank 600`*as.numeric(dilutionFactor)) %>%
                  group_by(strain, RpoS) %>%
                  summarise(meanDensity = mean(trueDensity))


combinedData <-left_join(OD420data, meanCellDensities, by=c("strain", "RpoS")) %>%
                select(`Blank 420`, experimenter, date, strain, RpoS, dilutionFactor, volumeAssayed, rea
                filter(strain != "blank") %>%
                rename(A420 = `Blank 420`) %>%
                type_convert() %>%
                mutate(`Miller Units` = 1000*A420*dilutionFactor/(meanDensity*volumeAssayed*reactionTim
```

```
## Parsed with column specification:
## cols(
##   experimenter = col_character(),
##   date = col_date(format = ""),
##   strain = col_character(),
##   RpoS = col_integer(),
##   dilutionFactor = col_integer(),
##   volumeAssayed = col_double(),
##   reactionTime = col_integer()
## )
```

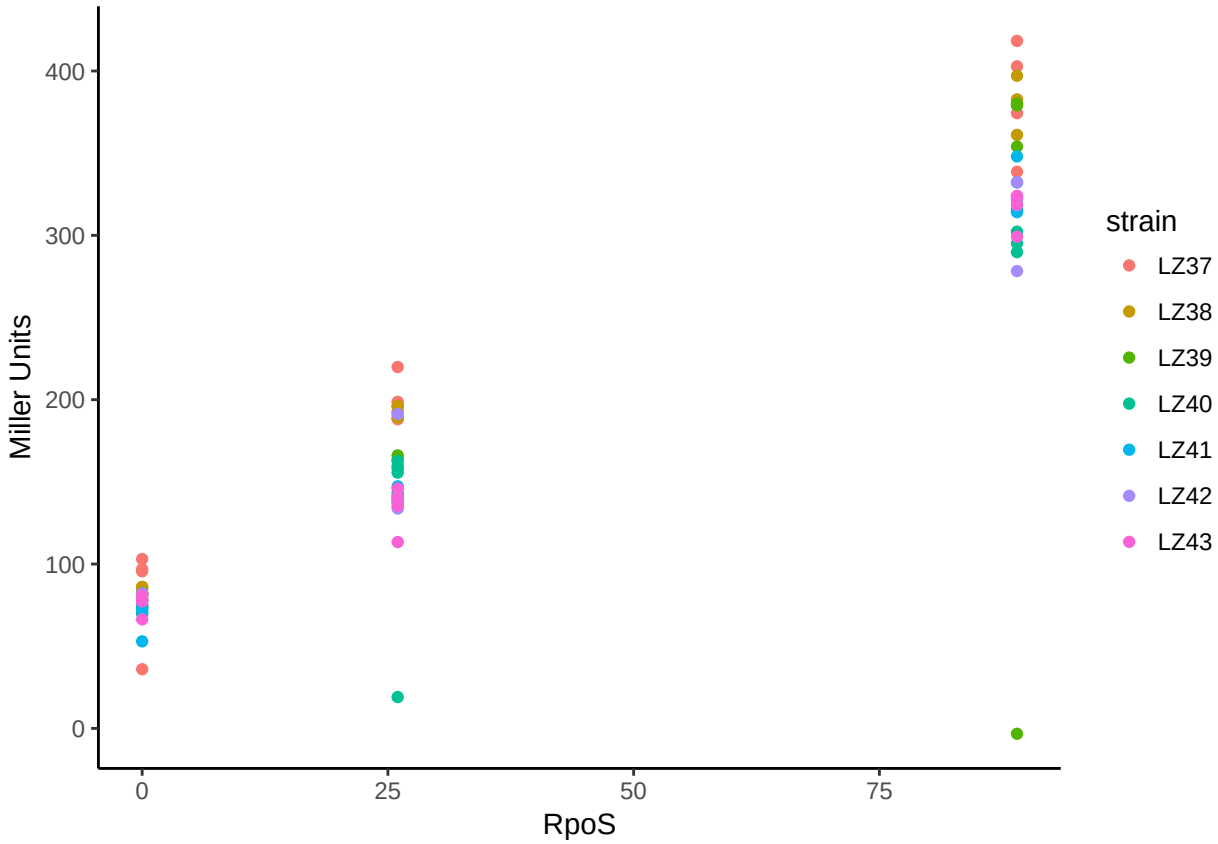
Plot the data

Two plots are shown, one using facets to create a plot for each strain, the other using colors to plot all replicates of various strains on a single plot. You don't need to use both in presentations, written documents, etc. Use the one that you think works best for your purpose.

```
ggplot(combinedData) + geom_point(aes(x=RpoS, y=`Miller Units`)) + facet_grid(.~strain)
```




```
ggplot(combinedData) + geom_point(aes(x=RpoS, y=`Miller Units`, color=strain))
```



Calculate and save the means at each concentration.

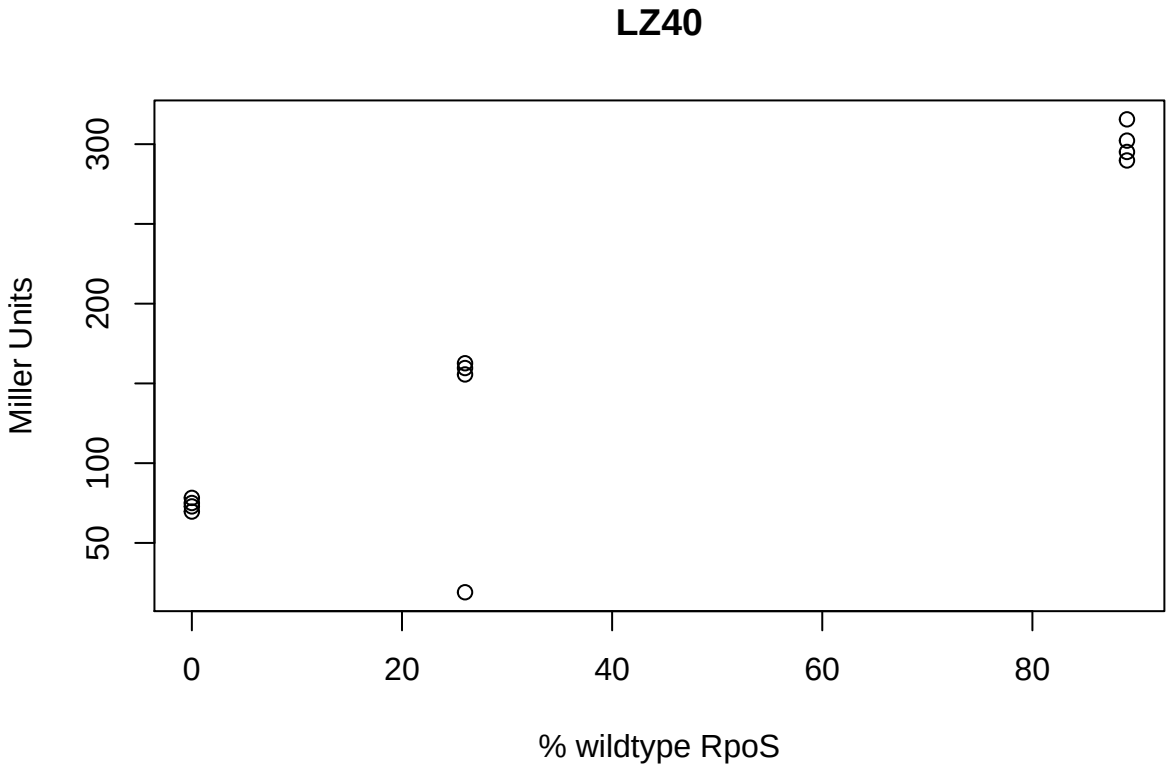
These are the replciates we use for actual calcualations of sensitivity

```
expressionLevels <- meanExpressionOneReplicate(fileName)
```

```
## Parsed with column specification:
## cols(
##   experimenter = col_character(),
##   date = col_date(format = ""),
##   strain = col_character(),
##   RpoS = col_integer(),
##   dilutionFactor = col_integer(),
##   volumeAssayed = col_double(),
##   reactionTime = col_integer()
## )
```

Plot individual strains

```
strainName <- "LZ40"
singlestrainMiller <- combinedData$`Miller Units`[combinedData$strain == strainName]
singlestrainRpoS <- combinedData$RpoS[combinedData$strain == strainName]


plot(singlestrainMiller ~ singlestrainRpoS, xlab = "% wildtype RpoS", ylab = "Miller Units", main = str
```


LZ40
Miller Units
50
100
200
300
0
20
40
60
80
% wildtype RpoS