

Simple document

I'm an R Markdown document!

Homework 1

Here's a **code chunk** that samples from a *normal distribution*:

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.2      v purrr  0.3.4
## v tibble  3.0.3      v dplyr  1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

hw_df =
  tibble(
    data1 = rnorm(10),
    samp_gt_0 = data1 > 0,
    char_vec = c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j"),
    factor_vec = factor(c("low", "low", "low", "mod", "mod", "mod", "high", "high", "high", "high"))
  )

mean(pull(hw_df, data1))

## [1] 0.174865

mean(pull(hw_df, samp_gt_0))

## [1] 0.6

mean(pull(hw_df, char_vec))

## Warning in mean.default(pull(hw_df, char_vec)): argument is not numeric or
## logical: returning NA

## [1] NA

mean(pull(hw_df, factor_vec))

## Warning in mean.default(pull(hw_df, factor_vec)): argument is not numeric or
## logical: returning NA

## [1] NA
```

We can take the mean of numbers and logical, but cannot compute the mean of vectors or factors.

```

as.numeric(pull(hw_df, data1))

## [1] 0.76471243 -1.27972936 0.08629515 1.16380516 0.10570419 1.02247736
## [7] -0.70600144 -0.53546903 -0.19889590 1.32575119

as.numeric(pull(hw_df, samp_gt_0))

## [1] 1 0 1 1 1 1 0 0 0 1

as.numeric(pull(hw_df, char_vec))

## Warning: NAs introduced by coercion
## [1] NA NA NA NA NA NA NA NA NA NA

as.numeric(pull(hw_df, factor_vec))

## [1] 2 2 2 3 3 3 1 1 1 1

Numeric, logical and factor can be converted to numeric, but vector cannot.

as.numeric(pull(hw_df, samp_gt_0)) * pull(hw_df, data1)

## [1] 0.76471243 0.00000000 0.08629515 1.16380516 0.10570419 1.02247736
## [7] 0.00000000 0.00000000 0.00000000 1.32575119

as.factor(pull(hw_df, samp_gt_0)) * pull(hw_df, data1)

## Warning in Ops.factor(as.factor(pull(hw_df, samp_gt_0)), pull(hw_df, data1)):
## '*' not meaningful for factors
## [1] NA NA NA NA NA NA NA NA NA NA

as.numeric(as.factor(pull(hw_df, samp_gt_0))) * pull(hw_df, data1)

## [1] 1.5294249 -1.2797294 0.1725903 2.3276103 0.2114084 2.0449547
## [7] -0.7060014 -0.5354690 -0.1988959 2.6515024

```

Homework 2

```

data("penguins", package = "palmerpenguins")
head(penguins)

## # A tibble: 6 x 8
##   species island bill_length_mm bill_depth_mm flipper_length~ body_mass_g sex
##   <fct>   <fct>         <dbl>         <dbl>         <int>     <int> <fct>
## 1 Adelie  Torge~           39.1           18.7           181       3750 male
## 2 Adelie  Torge~           39.5           17.4           186       3800 fema~
## 3 Adelie  Torge~           40.3            18           195       3250 fema~
## 4 Adelie  Torge~           NA            NA            NA         NA <NA>
## 5 Adelie  Torge~           36.7           19.3           193       3450 fema~
## 6 Adelie  Torge~           39.3           20.6           190       3650 male
## # ... with 1 more variable: year <int>

nrow(penguins)

## [1] 344

```

```
ncol(penguins)
```

```
## [1] 8
```

```
mean(pull(penguins, flipper_length_mm), na.rm = TRUE)
```

```
## [1] 200.9152
```

This dataset contains 344 penguins and 8 variables: species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, bill_depth_mm, body_mass_g, sex, year. The mean of flipper length is 200.9152047.

```
plot_df = tibble(  
  y = pull(penguins, flipper_length_mm),  
  x = pull(penguins, bill_length_mm)  
  
ggplot(plot_df, aes(x = x, y = y, color = pull(penguins, species))) + geom_point()
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

