

TA trial script

Mads Hartman (s184284)

2023-08-13

loading and cleaning data

```
data('gravier', package = 'datamicroarray')

clean_grav <- mutate(as_tibble(pluck(gravier,"x")),
                     y = pluck(gravier,"y")) %>%
  relocate(y) %>%
  rename( outcome = y) %>%
  mutate( outcome = case_when(
    outcome == "good" ~ 0,
    outcome == "poor" ~ 1))
```

transforming to long, and sampling 100

```
clean_grav_long <-
  clean_grav %>%
  pivot_longer(cols = starts_with("g"),
               names_to = "gene",
               values_to = "log2_expr_level") %>%
  group_by(gene) %>%
  nest() %>%
  ungroup() %>%
  sample_n(100)
```

fitting model and getting the results

```
results <- clean_grav_long %>%
  mutate(model_results = map(data, ~ glm(outcome ~ log2_expr_level,
                                         data = ., family = binomial)),
         model_info = map(model_results,
                          tidy))
```

extracting information

```
results_final <- results %>%
  mutate(
    ci_lower = map_dbl(model_results, ~ confint(.x)[2, 1]),
    ci_upper = map_dbl(model_results, ~ confint(.x)[2, 2]),
    pval_significant = model_info %>% map_dbl(~ .x[2,5] <= 0.05),
    model_coefficients = map(model_results, ~ pluck(.x)[1][1])
  ) %>%
  mutate(
    beta_1_est = map_dbl(model_coefficients, ~ pluck(.x$coefficients, "log2_expr_level"))
  ) %>%
  mutate(
    pval_significant = as.factor(pval_significant)
  )
```

removal of unneeded parts

```
plotting_results <- results_final %>%
  select(-data,
    -model_results,
    -model_info,
    -model_coefficients) %>%
  arrange(beta_1_est)
```

plotting

```
final_plot <- ggplot(data = plotting_results, aes(y = reorder(gene, beta_1_est, decreasing = T),
  x = beta_1_est,
  xmin = ci_lower,
  xmax = ci_upper,
  colour = pval_significant,
  height = 0.5)) +

  geom_point() +
  geom_errorbarh() +
  geom_vline(xintercept = 0, lty = "dashed") +
  xlab("Estimate") +
  ylab("") +
  theme(legend.position = "bottom")

final_plot
```

