Lecture 3

```
library (tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.2.1 ──
```

```
## ✔ ggplot2 3.0.0     ✔ purrr   0.2.5
## ✔ tibble  1.4.2     ✔ dplyr   0.7.7
## ✔ tidyr   0.8.1     ✔ stringr 1.3.1
## ✔ readr   1.1.1     ✔ forcats 0.3.0
```

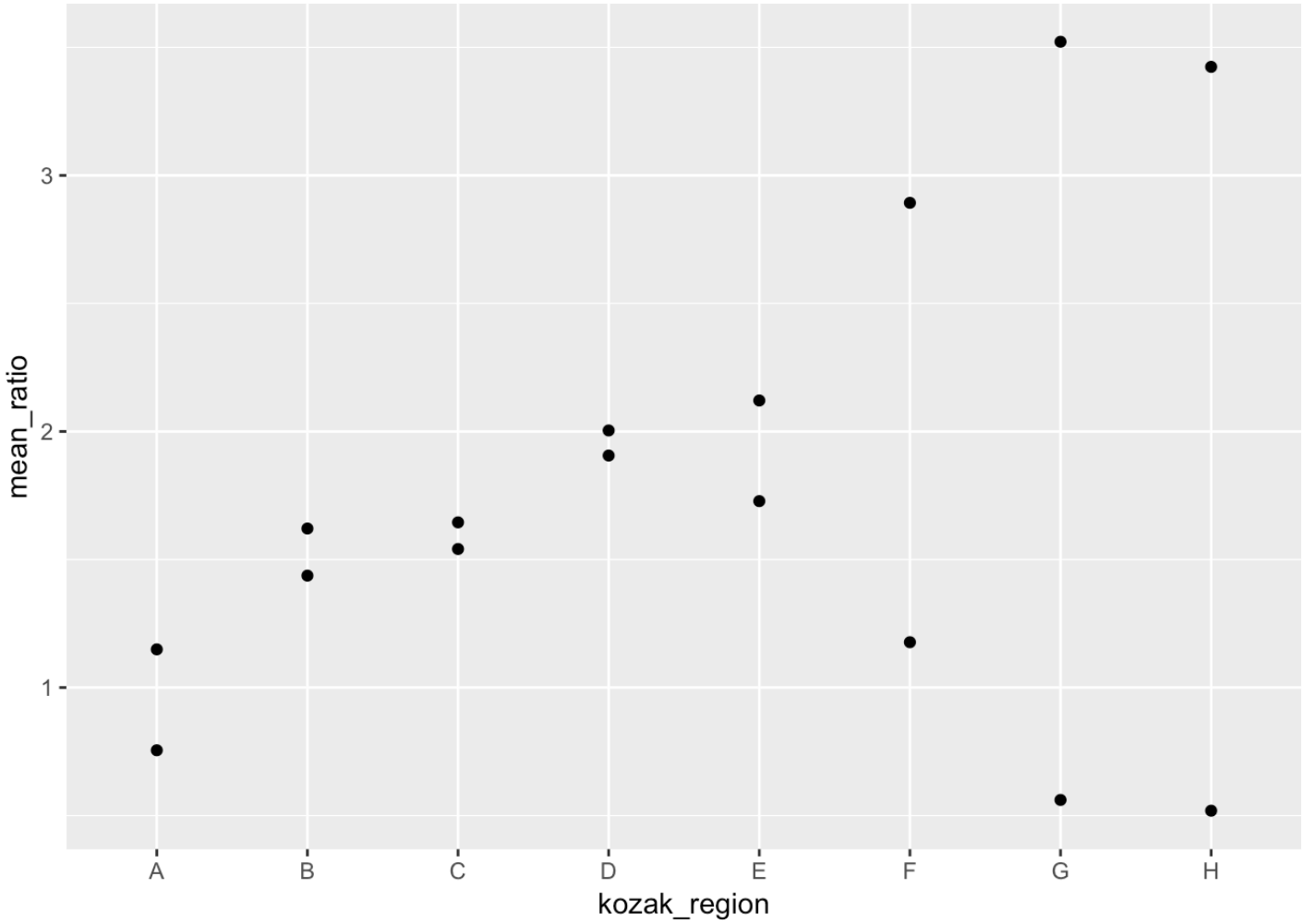
```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
data<-read_tsv("example_dataset_1.tsv") %>%
  print(data, n=5)
```

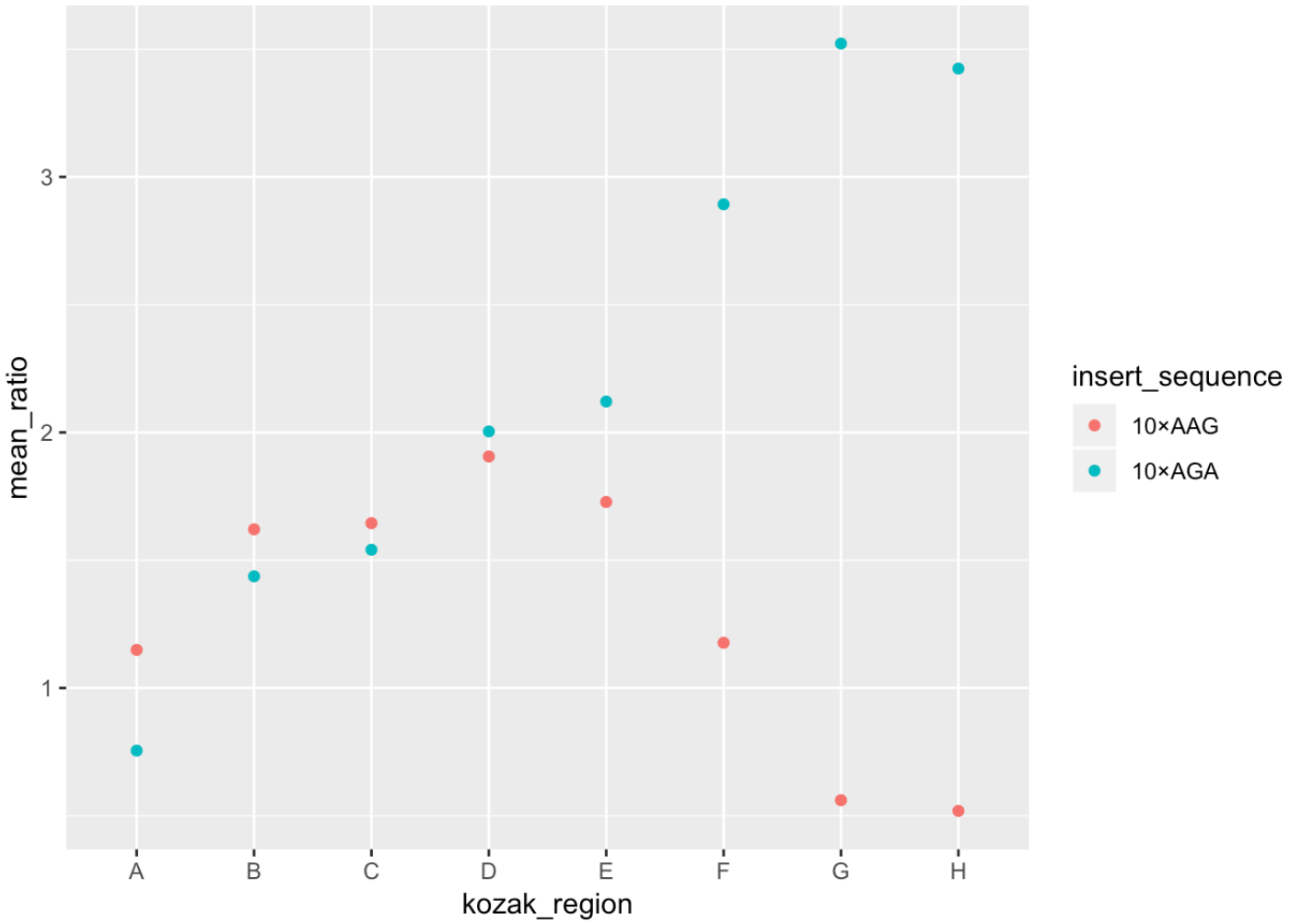
```
## Parsed with column specification:
## cols(
##   strain = col_character(),
##   mean_yfp = col_integer(),
##   mean_rfp = col_integer(),
##   mean_ratio = col_double(),
##   se_ratio = col_double(),
##   insert_sequence = col_character(),
##   kozak_region = col_character()
## )
```

```
## # A tibble: 16 x 7
##   stra… mean_yfp mean_rfp mean_ratio se_ratio insert_sequence kozak_region
##   <chr>    <int>    <int>      <dbl>    <dbl> <chr>           <chr>
## 1 schp…     1748    20754      0.755    0.066 10×AGA          A
## 2 schp…     3294    20585      1.44     0.021 10×AGA          B
## 3 schp…     3535    20593      1.54     0.018 10×AGA          C
## 4 schp…     4658    20860      2.00     0.021 10×AGA          D
## 5 schp…     5000    21171      2.12     0.023 10×AGA          E
## # ... with 11 more rows
```

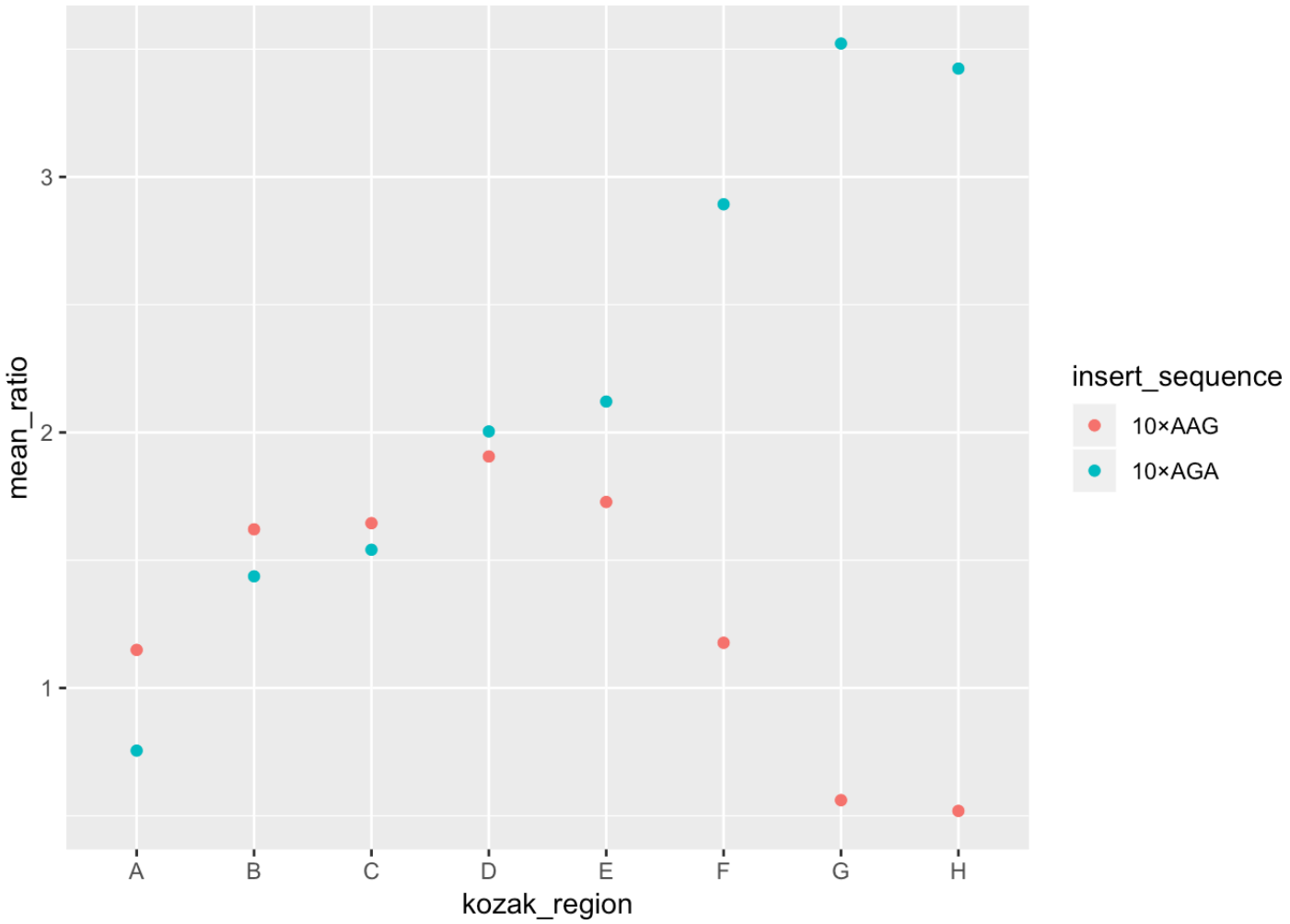
```
ggplot(data, aes(x = kozak_region,
                 y = mean_ratio)) +
  geom_point()
```



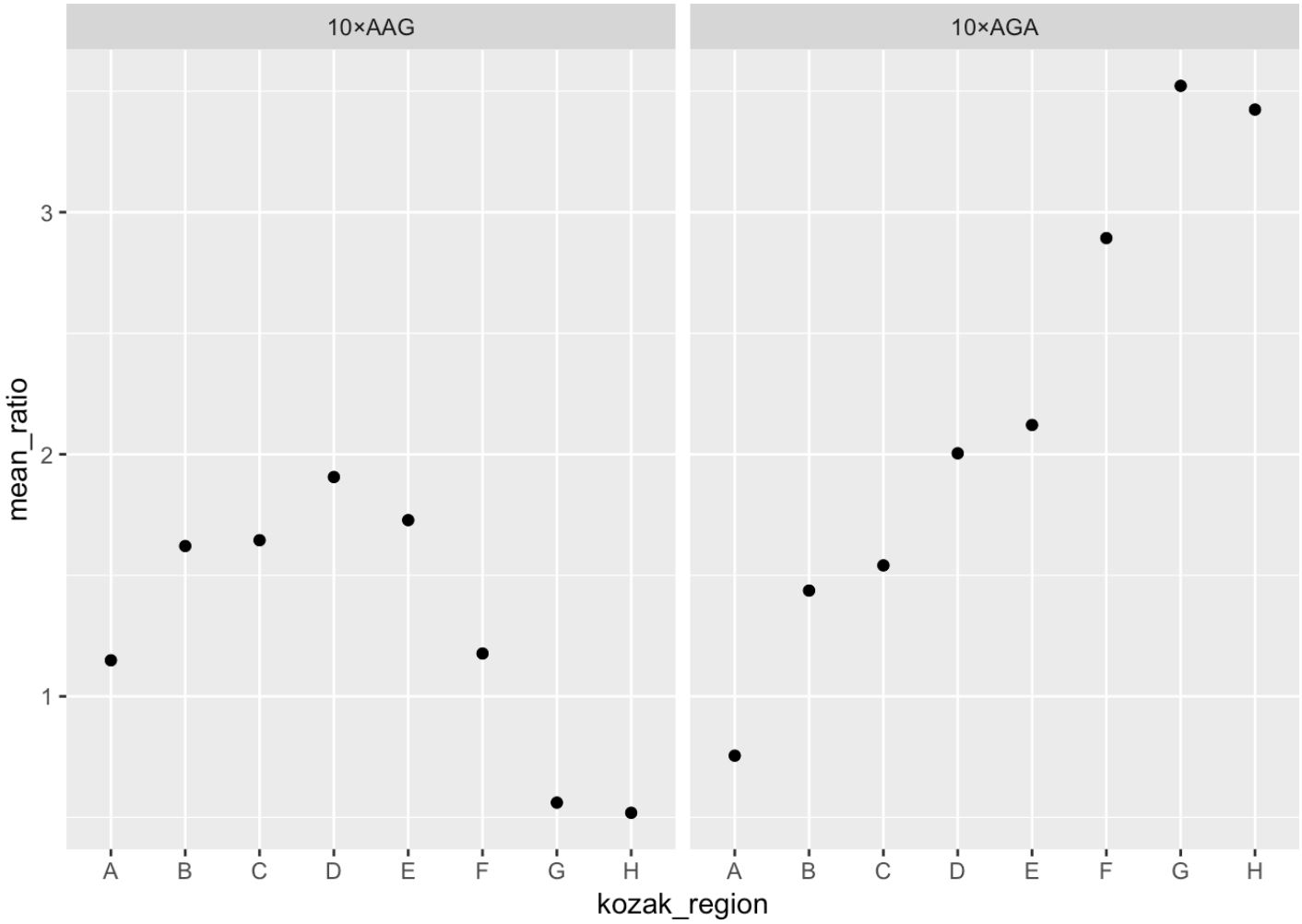

```
ggplot(data, aes(x = kozak_region,
                 y = mean_ratio,
                 color = insert_sequence,
                 group = insert_sequence)) +
  geom_point()
```




```
ggplot(data, aes(x = kozak_region,
                 y = mean_ratio,
                 color = insert_sequence,
                 group = insert_sequence)) +
  geom_point() +
  geom_point()
```




```
ggplot(data, aes(x = kozak_region,
                 y = mean_ratio,
                 group = insert_sequence)) +
  geom_point() +
  geom_point() +
  facet_grid(~insert_sequence)
```




```
print (data, n=3)
```

```
## # A tibble: 16 x 7
##   stra… mean_yfp mean_rfp mean_ratio se_ratio insert_sequence kozak_region
##   <chr>    <int>    <int>      <dbl>    <dbl> <chr>           <chr>
## 1 schp…     1748    20754      0.755    0.066 10×AGA          A
## 2 schp…     3294    20585      1.44     0.021 10×AGA          B
## 3 schp…     3535    20593      1.54     0.018 10×AGA          C
## # ... with 13 more rows
```

```
data %>%
  print(n=3)
```

```
## # A tibble: 16 x 7
##   stra… mean_yfp mean_rfp mean_ratio se_ratio insert_sequence kozak_region
##   <chr>    <int>    <int>      <dbl>    <dbl> <chr>           <chr>
## 1 schp…     1748    20754      0.755    0.066 10×AGA          A
## 2 schp…     3294    20585      1.44     0.021 10×AGA          B
## 3 schp…     3535    20593      1.54     0.018 10×AGA          C
## # ... with 13 more rows
```

```
data %>%
  print(n=2)
```

```
## # A tibble: 16 x 7
##   stra… mean_yfp mean_rfp mean_ratio se_ratio insert_sequence kozak_region
##   <chr>    <int>    <int>      <dbl>    <dbl> <chr>           <chr>
## 1 schp…     1748    20754      0.755    0.066 10×AGA          A
## 2 schp…     3294    20585      1.44     0.021 10×AGA          B
## # ... with 14 more rows
```

```
data %>%
  select(strain, mean_ratio, insert_sequence, kozak_region) %>%
  print(n=2)
```

```
## # A tibble: 16 x 4
##   strain  mean_ratio insert_sequence kozak_region
##   <chr>        <dbl> <chr>           <chr>
## 1 schp688      0.755 10×AGA          A
## 2 schp684      1.44  10×AGA          B
## # ... with 14 more rows
```

```
data %>%
  filter(kozak_region == "A")
```

```
## # A tibble: 2 x 7
##   stra… mean_yfp mean_rfp mean_ratio se_ratio insert_sequence kozak_region
##   <chr>    <int>    <int>      <dbl>    <dbl> <chr>           <chr>
## 1 schp…     1748    20754      0.755    0.066 10×AGA          A
## 2 schp…     2528    19906      1.15     0.056 10×AAG          A
```

```
data %>%
  filter(kozak_region == "A", insert_sequence == "10xAGA")
```

```
## # A tibble: 0 x 7
## # ... with 7 variables: strain <chr>, mean_yfp <int>, mean_rfp <int>,
## #   mean_ratio <dbl>, se_ratio <dbl>, insert_sequence <chr>,
## #   kozak_region <chr>
```

```
data %>%
  filter(kozak_region == "A") %>%
  filter(insert_sequence == "10xAGA")
```

```
## # A tibble: 0 x 7
## # ... with 7 variables: strain <chr>, mean_yfp <int>, mean_rfp <int>,
## #   mean_ratio <dbl>, se_ratio <dbl>, insert_sequence <chr>,
## #   kozak_region <chr>
```

```
data %>%
  arrange(mean_ratio)
```

```
## # A tibble: 16 x 7
##    strain mean_yfp mean_rfp mean_ratio se_ratio insert_sequence
##    <chr>     <int>    <int>      <dbl>    <dbl> <chr>
##  1 schp6…     1117    19377      0.519    0.01  10×AAG
##  2 schp6…     1270    20316      0.561    0.004 10×AAG
##  3 schp6…     1748    20754      0.755    0.066 10×AGA
##  4 schp6…     2528    19906      1.15     0.056 10×AAG
##  5 schp6…     2657    20223      1.18     0.048 10×AAG
##  6 schp6…     3294    20585      1.44     0.021 10×AGA
##  7 schp6…     3535    20593      1.54     0.018 10×AGA
##  8 schp6…     3687    20438      1.62     0.036 10×AAG
##  9 schp6…     3705    20227      1.64     0.021 10×AAG
## 10 schp6…     3967    20604      1.73     0.03  10×AAG
## 11 schp6…     4378    20630      1.91     0.01  10×AAG
## 12 schp6…     4658    20860      2.00     0.021 10×AGA
## 13 schp6…     5000    21171      2.12     0.023 10×AGA
## 14 schp6…     7379    22956      2.89     0.05  10×AGA
## 15 schp6…     8693    22649      3.42     0.125 10×AGA
## 16 schp6…     9365    23866      3.52     0.11  10×AGA
## # ... with 1 more variable: kozak_region <chr>
```