Comp 152-SBR: Statistical Bioinformatics in R

Fall 2021
T,R 10:30-11:45
summarize and group_by

Summarize creates a one-line summary of the whole tibble. Typically used after group_by to summarize groups:

```r
measles = group_by(filter(diseases, disease == "Measles"), year)
measles
```

# A tibble: 3,825 x 6
# Groups: year [75]

disease state year weeks_reporting count population
<fct> <fct> <dbl> <dbl> <dbl> <dbl>
1 Measles Alabama 1928 52 8843 2589923
2 Measles Alabama 1929 49 2959 2619131
3 Measles Alabama 1930 52 4156 2646248
4 Measles Alabama 1931 49 8934 2670818
summarize and group_by

Summarize creates a one-line summary of the whole tibble. Typically used after group_by to summarize groups:

```r
measles = group_by(filter(diseases, disease == "Measles"), year)
summarize(measles, total = sum(count))
```

# A tibble: 75 x 2
year  total
<dbl> <dbl>
1 1928  483337
2 1929  339061
3 1930  384597
4 1931  438435
...
summarize and group_by

Summarize creates a one-line summary of the whole tibble. Typically used after group_by to summarize groups:

```r
measles = group_by(filter(diseases, disease == "Measles"), year)
summarize(measles, total = sum(count))
plot(summarize(measles, total = sum(count)), type = "l")
```
Plotting groups with “facets”

Facets are different faces of the data, corresponding to “group_by” subsets.

Can plot each facet separately in ggplot.

commands: facet_grid (2d) and facet_wrap (1d)

for facet_grid, specify a formula,

vertical-var ~ horizontal-var

for facet_wrap with one variable, the formula is just ~ variable
meas.small = filter(diseases, disease == "Measles", state %in% c("California", "Massachusetts", "New York", "Ohio"))

ggplot(meas.small, aes(x = year, y = count)) + geom_line() + facet_wrap(~ state)
Facets on diabetes data

Use mutate and ifelse statements to create a new copy of diab called diabp that includes a categorical variable, pregcat, that maps the number of pregnancies into three possible groups:

<table>
<thead>
<tr>
<th>pregnancies</th>
<th>category</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>“none”</td>
</tr>
<tr>
<td>1-4</td>
<td>“few”</td>
</tr>
<tr>
<td>5-8</td>
<td>“many”</td>
</tr>
<tr>
<td>&gt;8</td>
<td>“lots”</td>
</tr>
</tbody>
</table>

Use facet_wrap to plot histograms of BMI (“mass”) by category.

If you have extra time, make the bin width 5, and see if you can figure out how to display the facets in the order listed above.