First figures: ggplot2

Eleutherodactylus Research Studio · Fall 2026

Tom Devitt

2026-09-02

Monday's payoff

Monday you learned to build a tidy table: columns are variables, rows are observations, cells hold one value.

Today that discipline pays off. **Every figure you make this semester**, from phenology curves to the figures in your team's manuscript, will be a tidy table plus a few lines of ggplot2.

ggplot2 works by **mapping columns to visual channels** (position, color, shape), so it needs the variables to be columns. A spreadsheet with dates in the headers can't be plotted until someone reshapes it. A tidy table plots immediately.

The grammar

Every ggplot has the same three-part recipe:

```
ggplot(data,                    # 1. a tidy table
       aes(x = ..., y = ...)) + # 2. columns -> visual channels
  geom_something()              # 3. how to draw it
```

That's the whole idea (the "grammar of graphics" of Wickham's ggplot2, in the framing we borrow from Claus Wilke's DSC 385). A scatterplot, a boxplot, and a phenology curve differ only in the `geom_`.

Today's practice table

Simulated data with realistic shapes and made-up numbers. Your teams will replace this with the real thing.

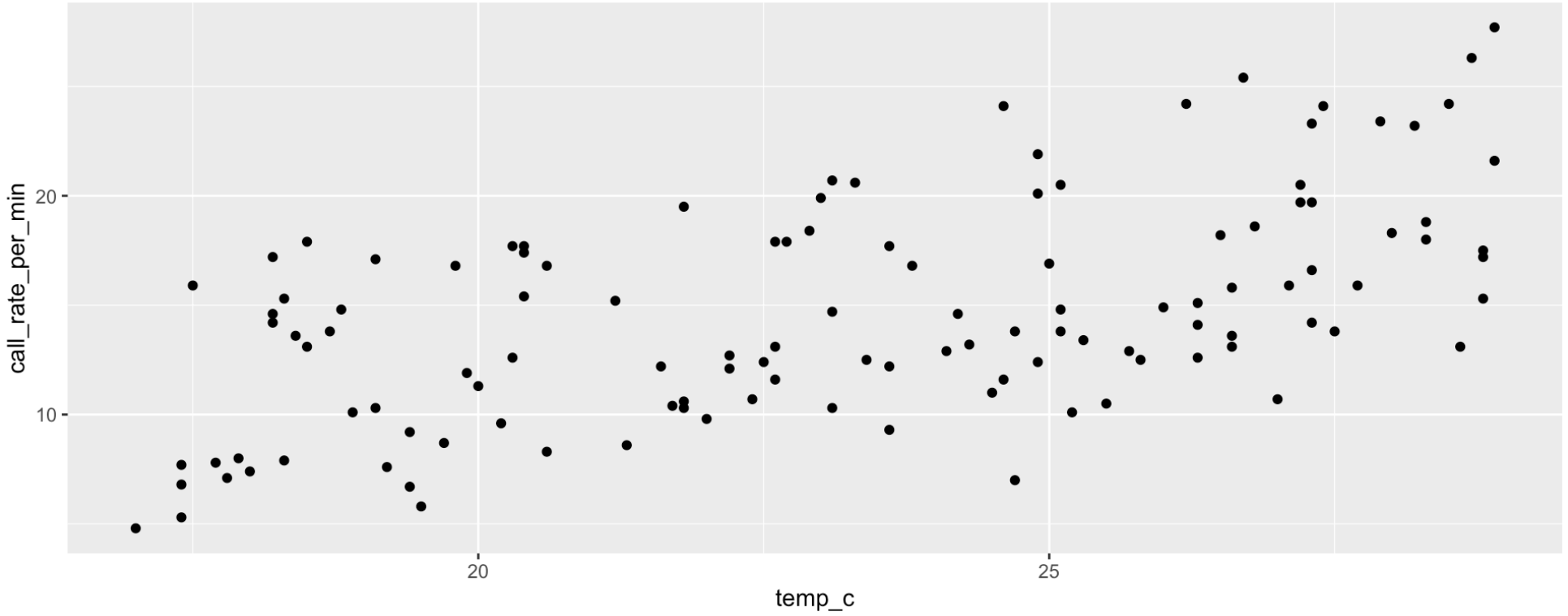
```r
set.seed(206)
calls <- tibble(
  species = rep(c("E. marnockii", "E. cystignathoides", "E. guttilatus"),
                each = 40),
  temp_c  = round(runif(120, 17, 29), 1),
  call_rate_per_min = round(rep(c(8, 14, 5), each = 40) +
                            0.9 * (temp_c - 17) + rnorm(120, 0, 1.8), 1),
  dominant_khz = round(rep(c(2.9, 4.1, 3.4), each = 40) +
                       0.04 * (temp_c - 17) + rnorm(120, 0, 0.12), 2)
)
head(calls, 4)
```

```
# A tibble: 4 × 4
  species      temp_c call_rate_per_min dominant_khz
  <chr>         <dbl>             <dbl>        <dbl>
1 E. marnockii   27.3              19.7         3.29
2 E. marnockii   22.6              13.1         3.25
3 E. marnockii   27.2              19.7         3.21
4 E. marnockii   18                 7.4         3.04
```

One row = one recorded male. Every variable is a column. Sound familiar?

First figure

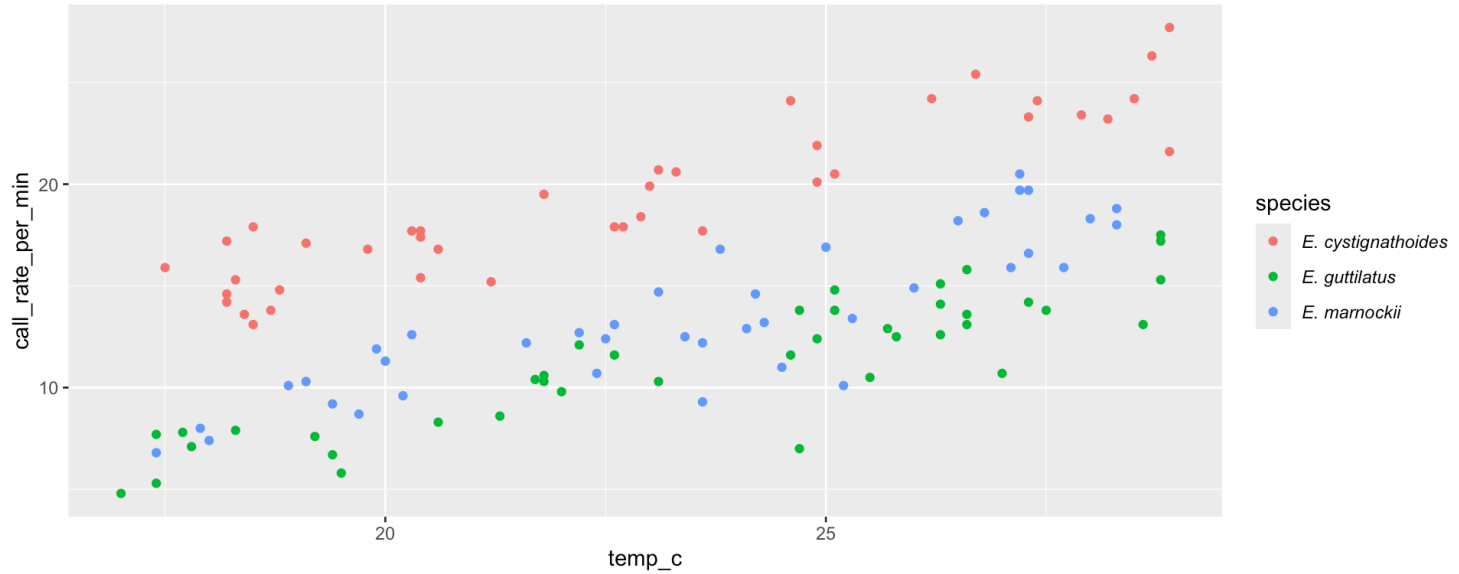
```
1 ggplot(calls, aes(x = temp_c, y = call_rate_per_min)) +
2   geom_point()
```



Data, mapping, geom, and the pattern is already visible. Warmer frogs call faster.

A third column, zero extra work

```
1 ggplot(calls, aes(x = temp_c, y = call_rate_per_min, color = species)) +
2   geom_point() +
3   theme(legend.text = element_text(face = "italic"))
```



`color = species` works **because species is a column**. This one-word change is what Monday's rules buy you.

(The `theme()` line italicizes the legend. Species names are italic **always**, even in figures. Copy the line for now; `theme()` gets its own treatment later.)

Mapped vs. set: the classic confusion

Inside `aes()`: mapped to data

```
1  aes(color = species)
```

Color *varies by* a column; you get a legend.

Outside `aes()`: set to a constant

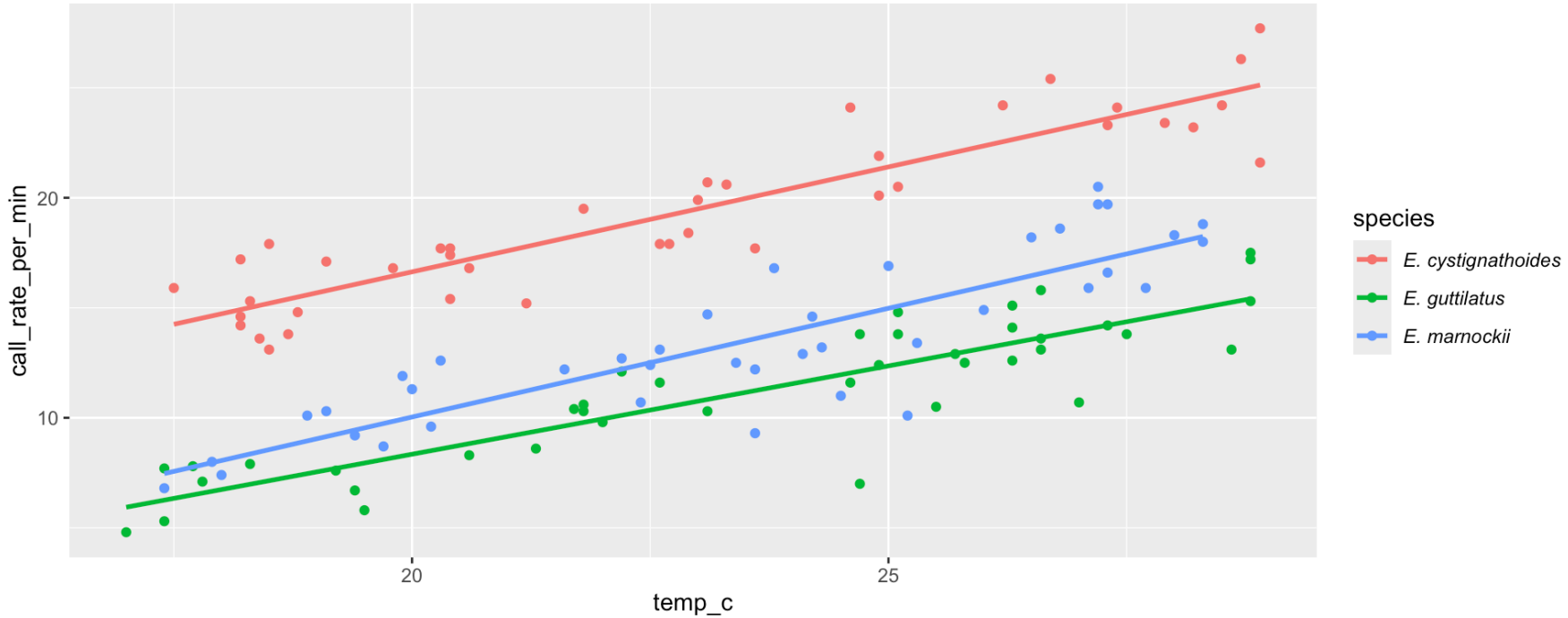
```
1  geom_point(color = "steelblue")
```

Everything is one color; no legend.

Everyone makes this error once: `aes(color = "blue")` maps color to the *word* “blue” and produces one salmon-pink legend entry named “blue.” When your colors look wrong, check which side of `aes()` you’re on.

Layers stack: adding a trend

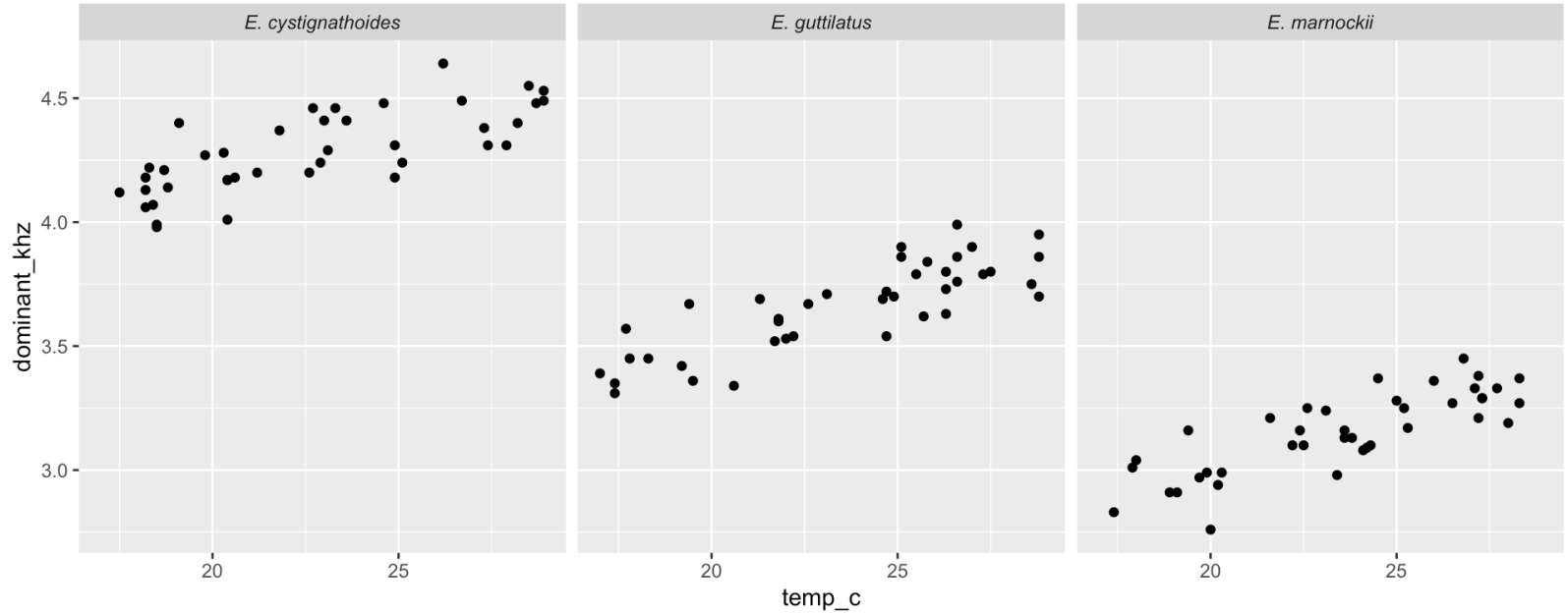
```
ggplot(calls, aes(x = temp_c, y = call_rate_per_min, color = species)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme(legend.text = element_text(face = "italic"))
```



Each + adds a layer. This figure (call rate vs. temperature by species) is exactly what the *marnockii* teams will build from the AudioMoth data, for real.

Facets: small multiples

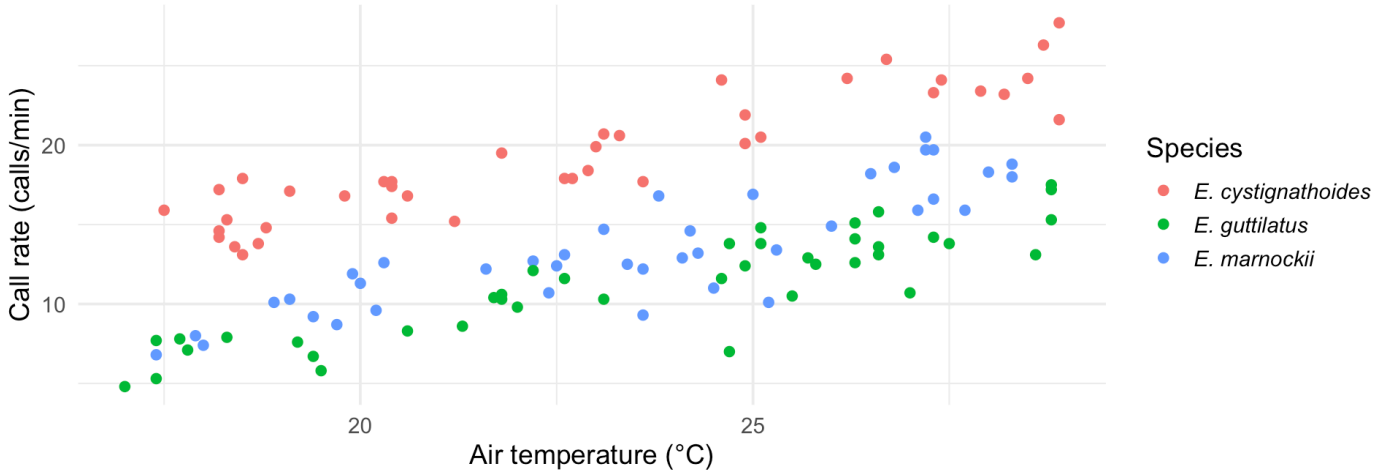
```
ggplot(calls, aes(x = temp_c, y = dominant_khz)) +
  geom_point() +
  facet_wrap(~ species) +
  theme(strip.text = element_text(face = "italic"))
```



One panel per group with identical axes, which is the honest way to compare species. It works, once again, because `species` is a column. (Facet labels are “strips”; same italics rule.)

Labels: the last 10% that makes it publishable

```
ggplot(calls, aes(x = temp_c, y = call_rate_per_min, color = species)) +
  geom_point() +
  labs(x = "Air temperature (°C)", y = "Call rate (calls/min)",
       color = "Species") +
  theme_minimal(base_size = 14) +
  theme(legend.text = element_text(face = "italic"))
```



Column names carry units for the computer (`temp_c`); `labs()` translates them for humans. No figure leaves the class without labeled axes, units, and italic species names.

Saving

```
ggsave("figures/call-rate-vs-temp.png", width = 7, height = 4.5, dpi = 300)
```

- Saves the **most recent** plot, at a size and resolution you chose and can repeat. No screenshots or dragged plot-pane corners
- Remember Monday: figures are *derived* products. The script that makes them reads the raw table; you can delete every figure and regenerate them all

Your turn: worksheet

On the EDU pod:

```
file.copy("/stor/work/FRI_Bio206_TD_Spring2026/04-ggplot2-worksheet.qmd", "~/")
```

Open it, work through the five tasks (they use today's simulated `calls` table), **render it, and submit the rendered file to Canvas** like you did with `hello.qmd`.

Finish early? The stretch task previews next week: distributions.

Assignment: find your species' voice

No class Monday (Labor Day). Instead, meet the frog you just claimed. **Due Wed Sep 9 on Canvas:** one rendered Quarto HTML containing:

0. *First*, skim the **terminology section of Köhler et al. 2017** (link on Resources), just the part that defines **call, note, and pulse**, plus its figures. Twenty minutes now and you'll know a call from a note for the rest of the semester
1. Your claimed species and its **species group** (projects page has the map)
2. The result of a **recording hunt**: search AmphibiaWeb, the Macaulay Library, Fonoteca Zoológica, and Google Scholar for a published or archived recording of its call. Link what you find. For many of our species **no public recording exists, and documenting that absence is a real result**. That gap is why this class exists
3. **2–3 sentences describing the call, using Köhler's terms where they fit**: one note or many? A tonal whistle, a chirp, or a pulsed trill? Repeated how often? If your species has no recording, describe the closest relative you can find (start within its species group)
4. **One figure** from today's worksheet table, fully polished: labels, units, italic species names

This week

- **Team sign-up closes TONIGHT, 11:59 pm** (Canvas > People > Groups). First come, first served, four per team
- **AmphibiaWeb species claiming also closes today** (link on Canvas)
- **No class Monday** (Labor Day); do the species-voice assignment instead
- **Wed Sep 9:** Visualization II (histograms and boxplots for summarizing call measurements), then bioacoustics begins: sound, spectrograms, and how to read a frog call (Köhler et al. 2017)