

ggplot – groups, facets

Professor Halterman

Michigan State University

PLS 397 Analyzing and Visualizing Data
Fall 2023

Check-in and logistics

Today's quick check-in:

<https://forms.gle/j9ahmPcaeiGF83Pz7>

This should take about 5 minutes.

Logistics:

- ▶ In-class exercises are due the **following morning** at 9am.
- ▶ There will be **no class** the Wednesday before Thanksgiving.

This week's reading

- ▶ Wednesday: 4.6, 4.7
- ▶ Following Monday: 5.2

Table of Contents

- 1 Reviewing what we've learned
- 2 Groups and facets
- 3 In class exercise
- 4 Composing multiple plots

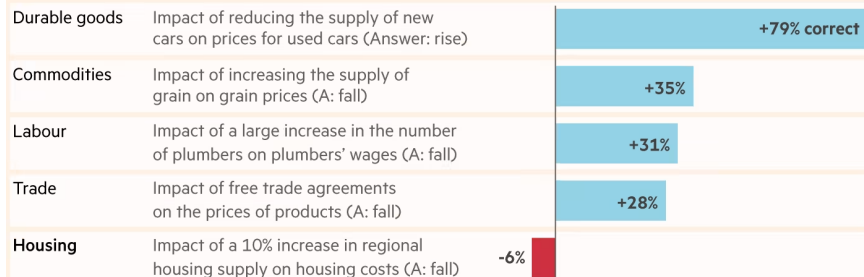
Housing costs in the US

- ▶ Housing costs are very high and increasing in the United States.
- ▶ It's clear that the issue is a lack of supply—relatively few houses get built, especially in larger cities.
- ▶ Debate: will building market rate housing decrease housing costs? Or do we need to build below market rate/subsidized housing to reduce prices?

Financial Times housing costs article

People are generally good at predicting how changes in supply impact prices — except when it comes to housing

Net correct minus incorrect answers to questions* about the impact of supply shocks



*See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4266459 for full question wordings

Source: *Folk Economics and the Persistence of Political Opposition to New Housing* (Nall et al., 2022)

FT graphic: John Burn-Murdoch / @jburnmurdoch

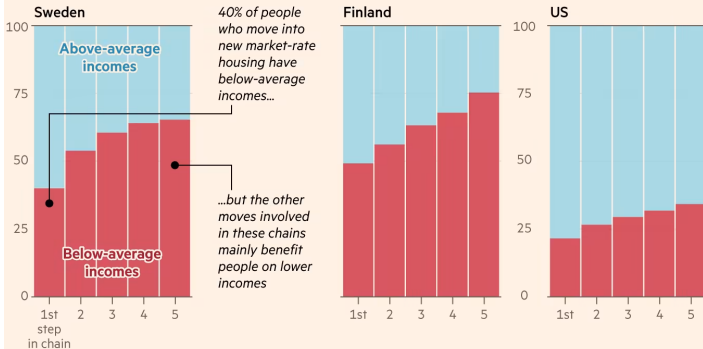
© FT

+ coord_flip()

Financial Times housing costs article

Most people who move into new market-rate housing are higher earners, but the chain of moves this triggers frees up homes for lower earners

Income breakdown of in-moving residents at each stage of the chain triggered by moves into new market-rate housing



Sources: City-wide effects of new housing supply: Evidence from moving chains (Bratu et al., 2023); Uppsala University Urban Lab; The Effect of New Market-Rate Housing Construction on the Low Income Housing Market (Mast, 2001)

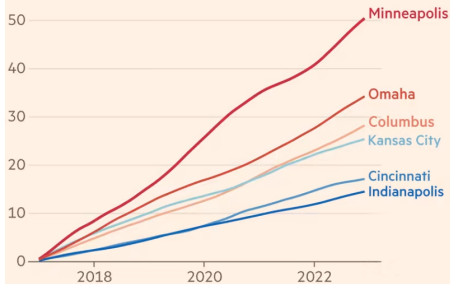
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT

+ position = "fill"

Financial Times housing costs article

Minneapolis has consistently built more housing than other Midwestern cities...

Cumulative new dwelling approvals per 1,000 people



*Rents deflated by average incomes

Sources: FT analysis of data from State of the Cities Data Systems, Apartment List, BLS, Census Bureau

FT graphic: John Burn-Murdoch / @jburnmurdoch

© FT

...and is reaping the rewards as rents fall relative to inflation

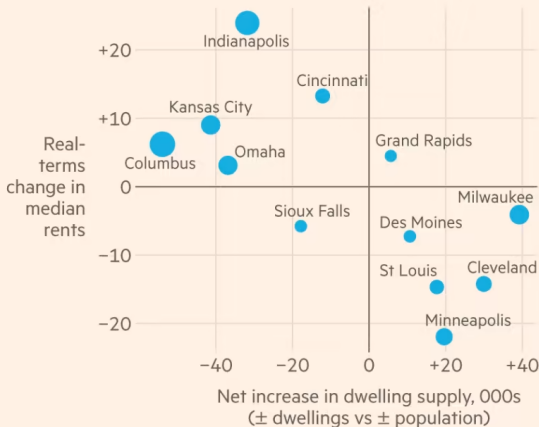
Real-terms change in median rent (Jan 2017 = 100)



Financial Times housing costs article

A similar pattern can be seen across the Midwest as a whole

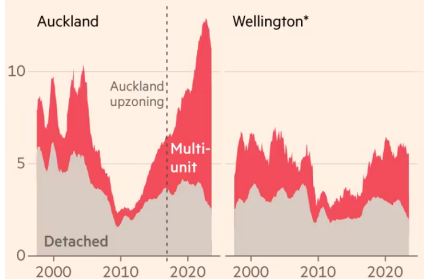
Net change in housing supply vs real-terms rent growth



Financial Times housing costs article

Upzoning in Auckland spurred a surge in high-density housebuilding...

New dwelling approvals per 1,000 people (rolling 1 yr avg)



*Excludes Lower Hutt, which upzoned in 2020

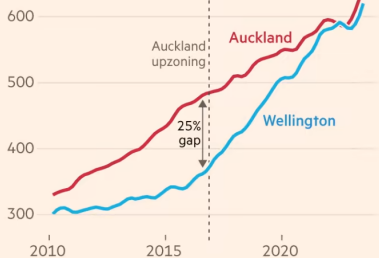
Sources: FT analysis of data from Stats NZ, New Zealand Tenancy Services, Reserve Bank of New Zealand. Based on prior work by Matthew Maltman

FT graphic: John Burn-Murdoch / @jburnmurdoch

© FT

...which slowed rent rises, erasing a 25% premium compared to Wellington...

Nominal median monthly rents (NZ\$), seasonally adjusted



Financial Times housing costs article

...and caused rents to stabilise after adjusting for inflation

Real-terms change in median rent (Nov 2016 = 100)

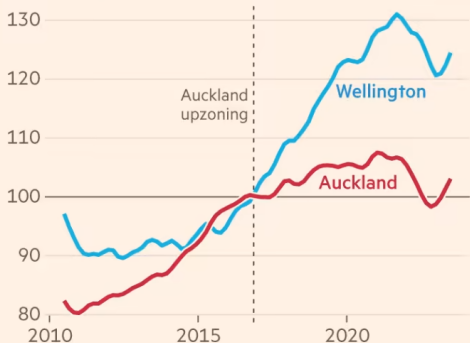


Table of Contents

1 Reviewing what we've learned

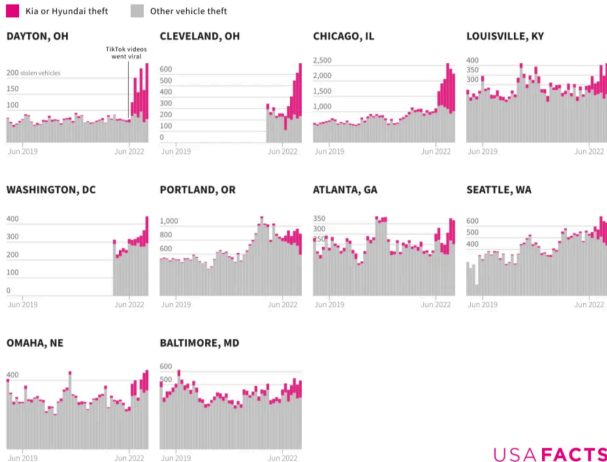
2 Groups and facets

3 In class exercise

4 Composing multiple plots

Small multiples

Thefts of **Kias and Hyundais** increased in 10 cities after videos demonstrating how to steal the cars went viral on TikTok.



Charts from USAFacts (usafacts.org/data-projects/car-thefts)

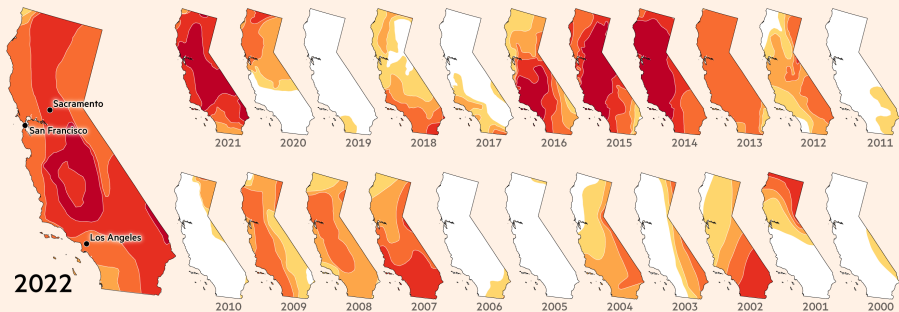
Data from police departments in Atlanta; Baltimore; Chicago; Cleveland; Dayton; Louisville Metro; Omaha; Portland; Seattle and Washington, DC

Small multiples

California has experienced drought conditions in July for the past 22 years

Drought level: Abnormally dry Moderate Severe Extreme Exceptional

Historical drought maps for week ending July 25 (+/- 3 days)



Source: US Drought Monitor
© FT

Cartography: Steven Bernard

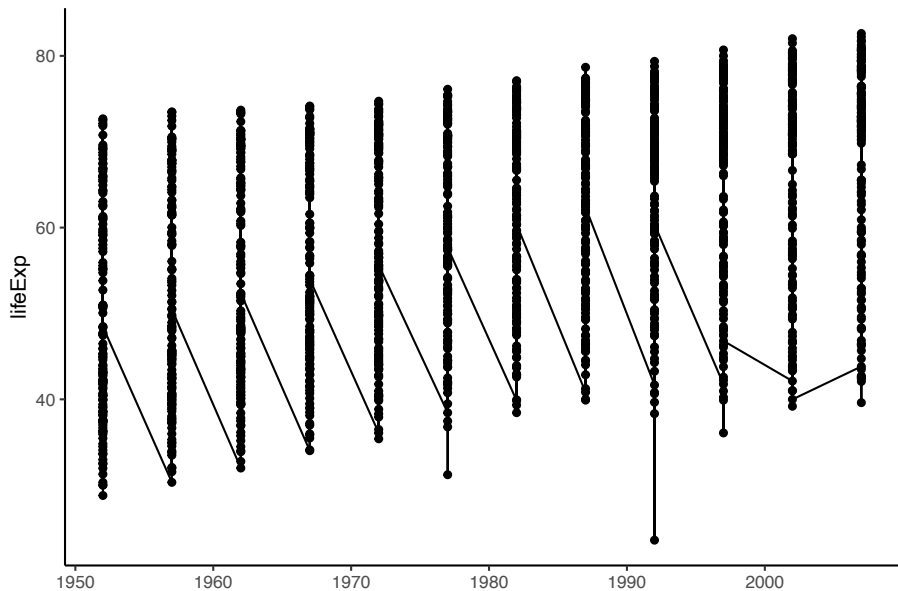
Small multiples/facets

- ▶ “Small multiples” (Tufte) or faceted plots are a very useful tool:
 - Avoid cluttered plots
 - Easy way to compare across **discrete** variables
 - Cognitively easy: same basic plot, repeated.
- ▶ How do we do this in ggplot?
- ▶ Step 1: make sure our aes call includes our facetting variable
- ▶ + `facet_wrap(~ variable)`
- ▶ Advanced use: if we have **two** facetting variables, we can use + `facet_grid(var1 ~ var2)`

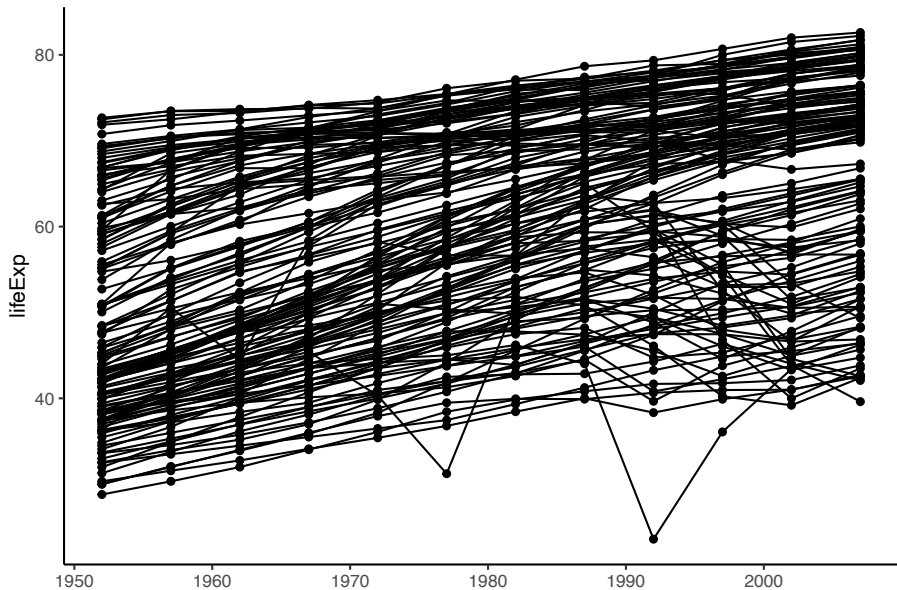
Let's start with a simple plot:

- ▶ Using the full gapminder dataset...
- ▶ ...make a line plot...
- ▶ ...with year as the x axis and life expectancy as the y axis

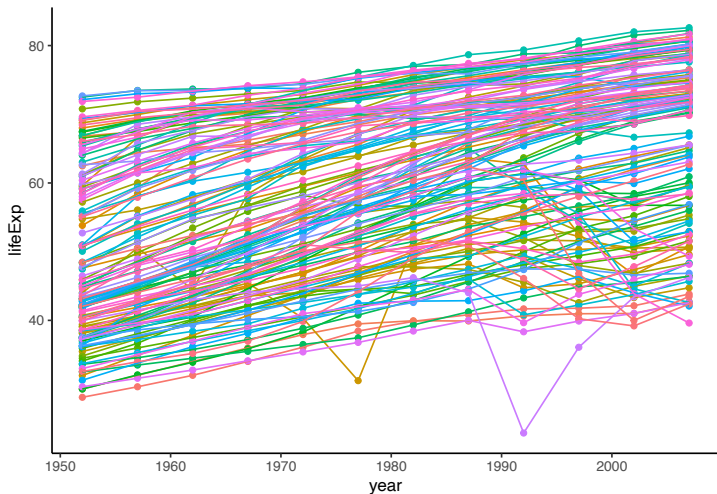
What's wrong?



Specifying country as the group

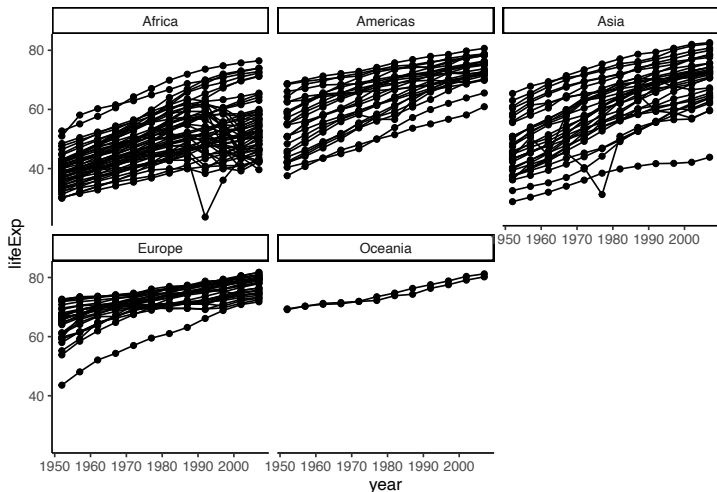


Setting the color also gives us a group

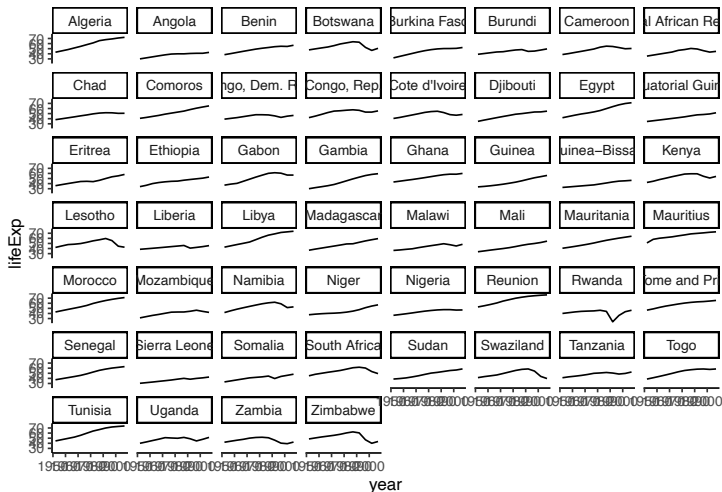


```
+ theme(legend.position = "none")
```

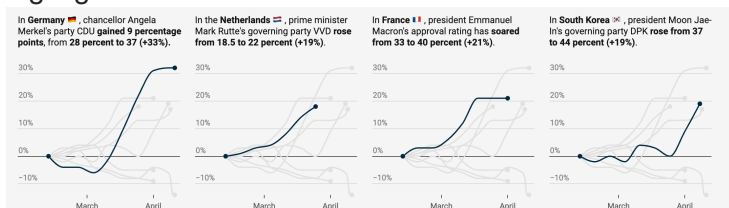
This is really busy! Can we facet by continent?



Show faceted line graphs for just Africa



- ▶ A powerful technique is to show all the data in each facet, but highlight each line in turn:

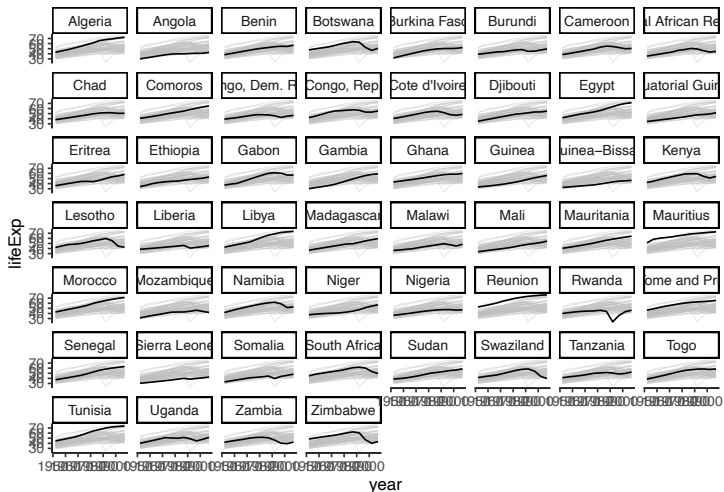


Source:

Datawrapper

- ▶ We can do this with `gghighlight`
- ▶ `+ gghighlight() + facet_wrap(...`

With highlighting



We can change the color, too

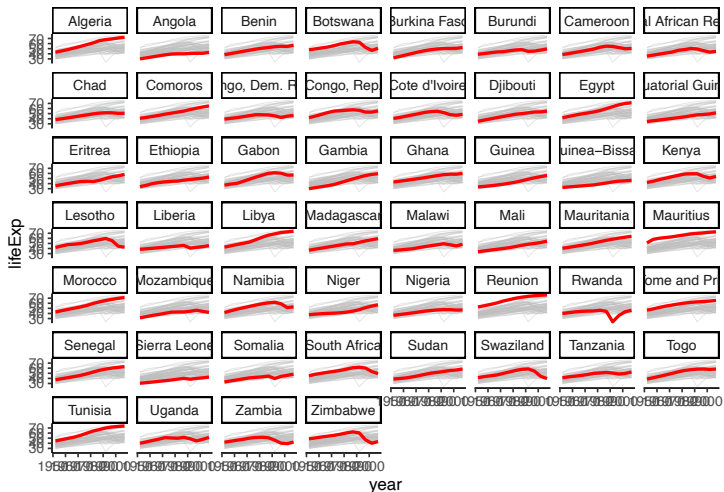


Table of Contents

1 Reviewing what we've learned

2 Groups and facets

3 In class exercise

4 Composing multiple plots

Polity dataset

- ▶ The Polity project codes countries' democracy scores on a -10 to +10 scale, with full autocracies being -10 and full democracies as +10
- ▶ Run the code to load the polity dataset and merge it with gapminder.

In-class exercise

Think of some questions you'd like to answer using the data, then make 2-3 plots to explore it. To get you started:

- ▶ What's the relationship between GDP and democracy?
- ▶ Do democracies have higher life expectancy?
- ▶ How has this changed over time?
- ▶ And by continent?
- ▶ Make sure to label and scale your plots!

Table of Contents

1 Reviewing what we've learned

2 Groups and facets

3 In class exercise

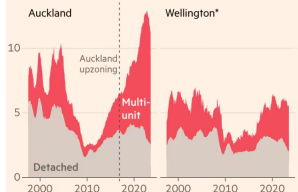
4 Composing multiple plots

Combining plots

- ▶ We've learned how to make **small multiples** using ggplot's facets
- ▶ But small multiples are the same plot, repeated by some other categorical variable.
- ▶ What if we want to put different plots side by side?

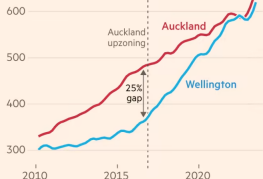
Upzoning in Auckland spurred a surge in high-density housebuilding...

New dwelling approvals per 1,000 people (rolling 1 yr avg)



...which slowed rent rises, erasing a 25% premium compared to Wellington...

Nominal median monthly rents (NZ\$), seasonally adjusted



*Excludes Lower Hutt, which upzoned in 2020

Sources: FT analysis of data from Stats NZ, New Zealand Tenancy Services, Reserve Bank of New Zealand. Based on prior work by Matthew Maltman

FT graphic: John Burn-Murdoch / @burnmurdoch

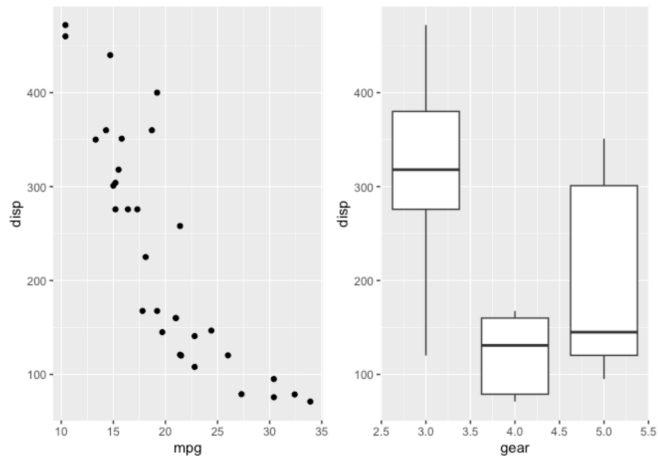
© FT

- ▶ Patchwork is an easy to use library for combining multiple ggplot figures
- ▶ Go ahead and install it: `install.packages('patchwork')`
- ▶ To use, we assign our plots to **objects**: `plot1 <- ggplot(...)`
- ▶ ...then we can use the + and / symbols to put plots next to each other or on top.
- ▶ Parentheses allow us to group plots together.

```
library(ggplot2)
library(patchwork)

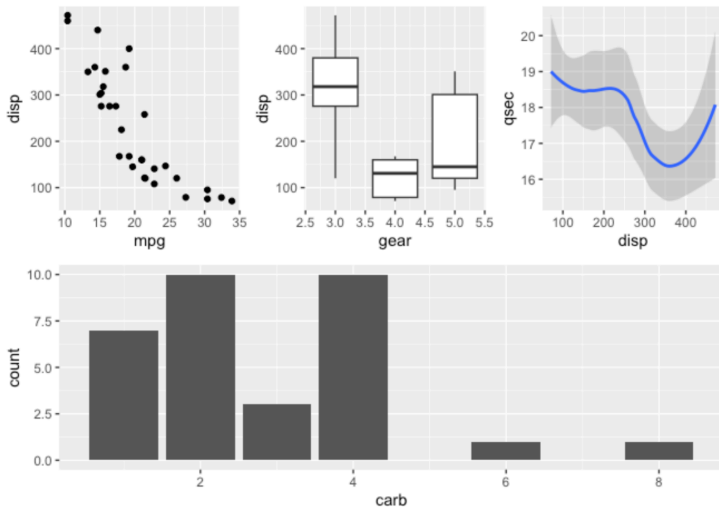
p1 <- ggplot(mtcars) + geom_point(aes(mpg, disp))
p2 <- ggplot(mtcars) + geom_boxplot(aes(gear, disp, group = gear))

p1 + p2
```



```
p3 <- ggplot(mtcars) + geom_smooth(aes(displ, qsec))
p4 <- ggplot(mtcars) + geom_bar(aes(carb))
```

(p1 | p2 | p3) /
p4



More patchwork options

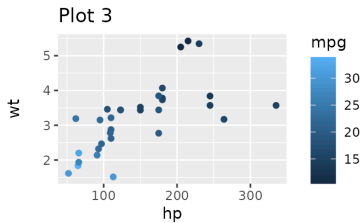
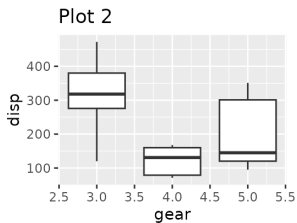
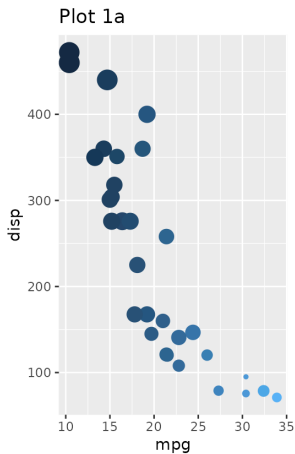
Patchwork gives you a lot of flexibility to do things like:

- ▶ Add spacing between plots
- ▶ Move the legend around
- ▶ Adding insets (smaller figures overlaid on top of a larger figure)
- ▶ Adding text panels
- ▶ Consolidating duplicate legends

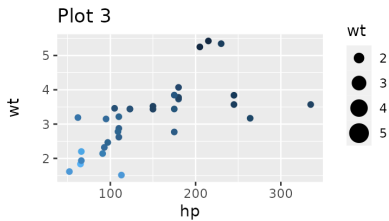
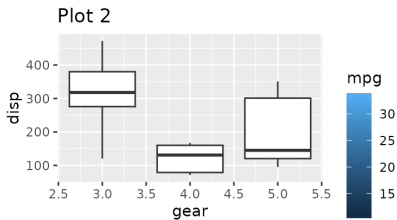
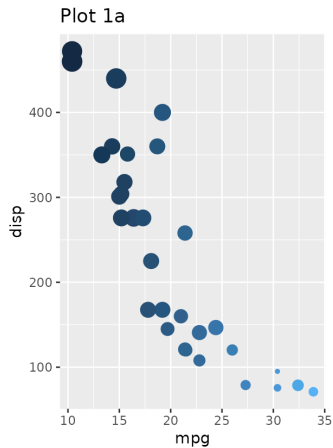
These will be very helpful when it comes to the midterm and final projects!

```
p1a <- ggplot(mtcars) +  
  geom_point(aes(mpg, disp, colour = mpg, size = wt)) +  
  ggtitle('Plot 1a')
```

p1a | (p2 / p3)



```
(p1a | (p2 / p3)) + plot_layout(guides = 'collect')
```



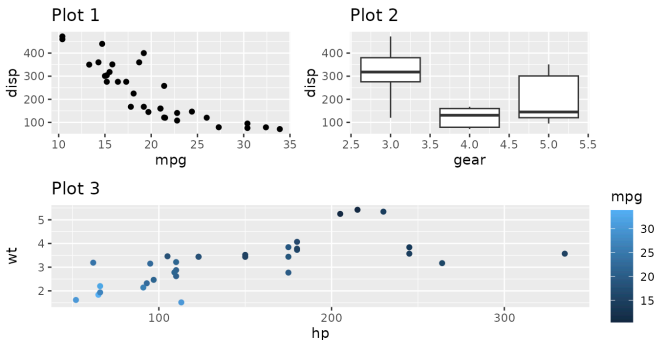
```

patchwork <- (p1 + p2) / p3
patchwork + plot_annotation(
  title = 'The surprising truth about mtcars',
  subtitle = 'These 3 plots will reveal yet-untold secrets about our beloved data-set',
  caption = 'Disclaimer: None of these plots are insightful'
)

```

The surprising truth about mtcars

These 3 plots will reveal yet-untold secrets about our beloved data-set



Disclaimer: None of these plots are insightful

Now combine your plots into a single patchwork plot and add an overall title.