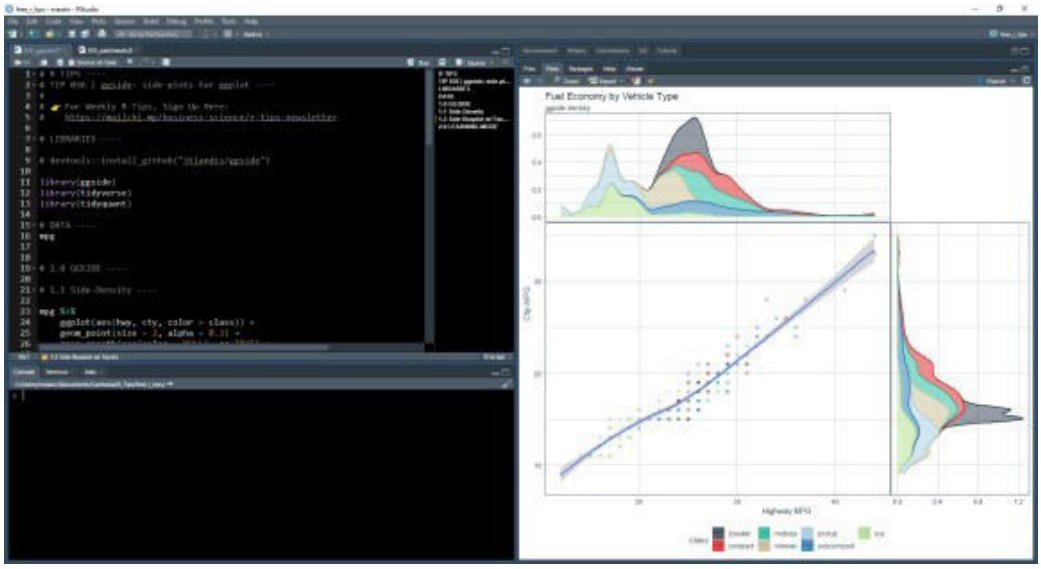
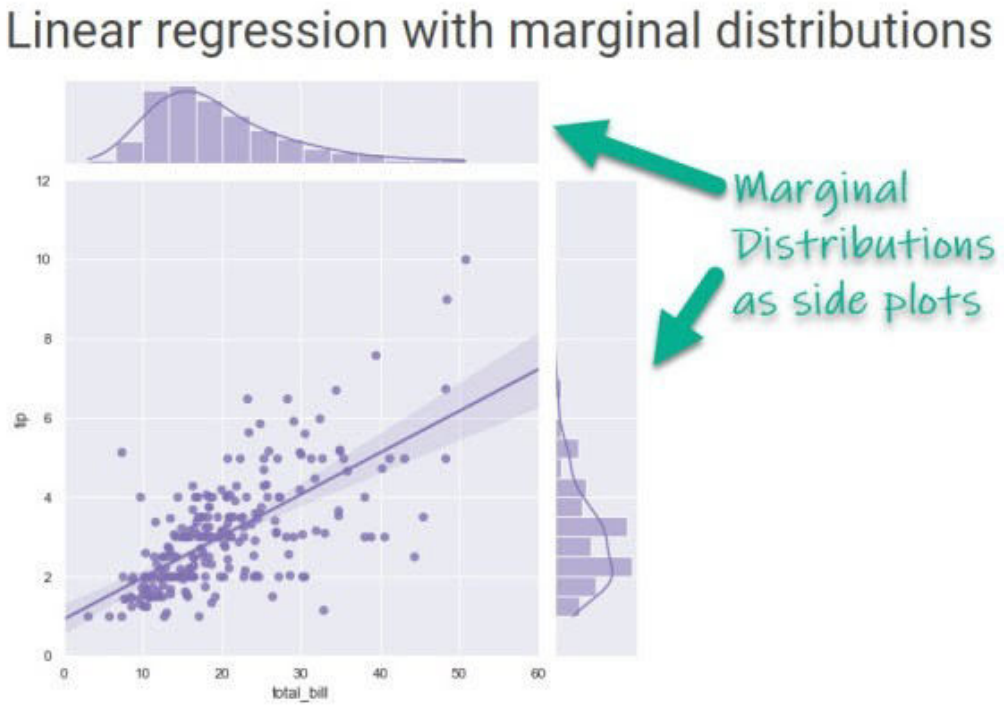
What are Marginal Distributions?

Marginal Distribution (Density) plots are a way to extend your numeric data with side plots that highlight the density (histogram or boxplots work too).



Linear Regression Marginal Distribution Side Plots.
One of two plots we're making today.

Marginal Distribution Plots were made popular with the seaborn jointplot() side-panels in Python. These add side plots that highlight distributions.



Seaborn's jointplot() makes a Linear Regression with Marginal Distributions.

How do we make them in ggplot2?

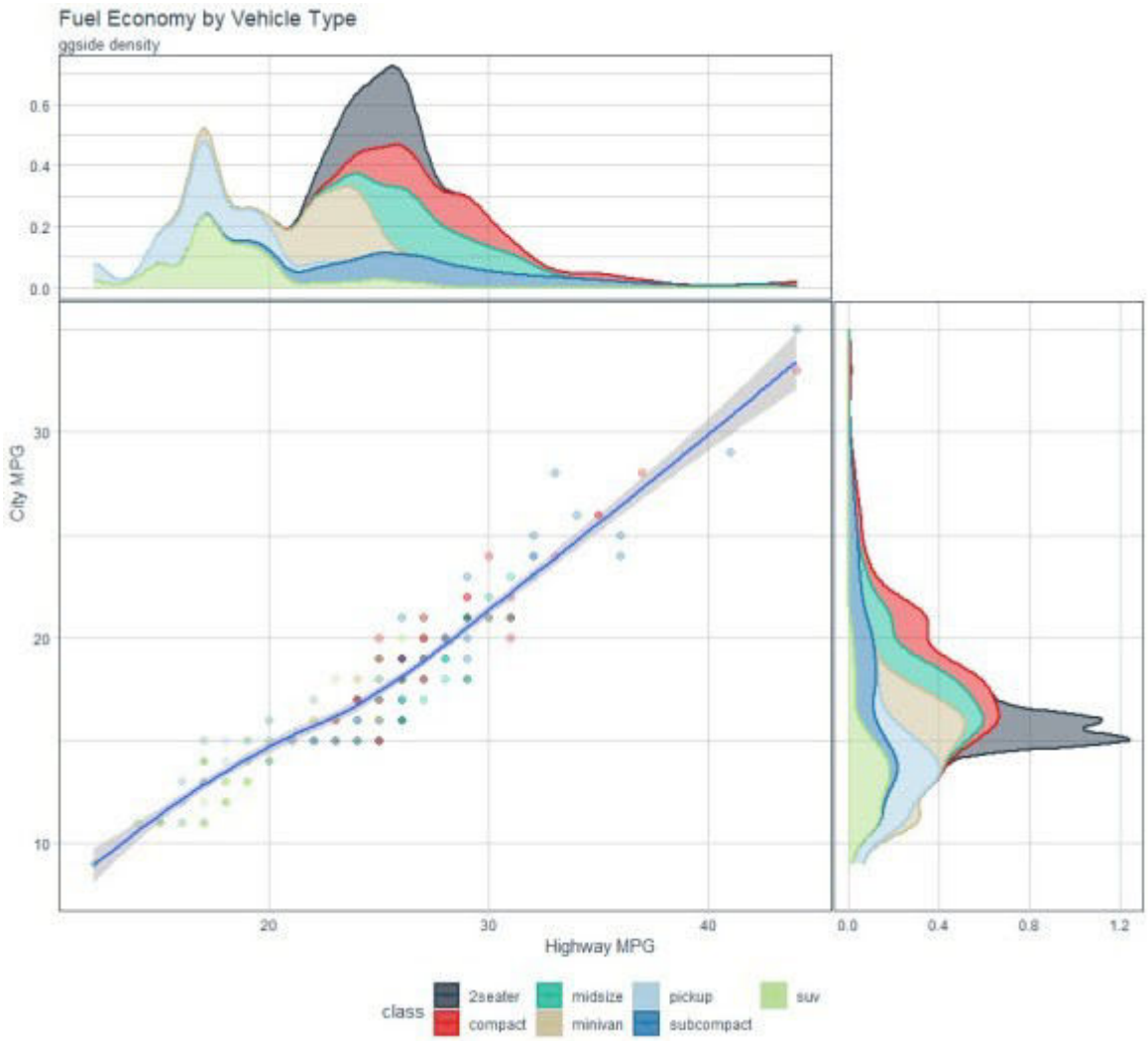
Marginal distributions can now be made in R using `ggside`, a new ggplot2 extension. You

can make linear regression with marginal distributions using histograms, densities, box plots, and more. Bonus – The side panels are super customizable for uncovering complex relationships.

Here are **two examples** of what you can (and will) do in this tutorial! 👇

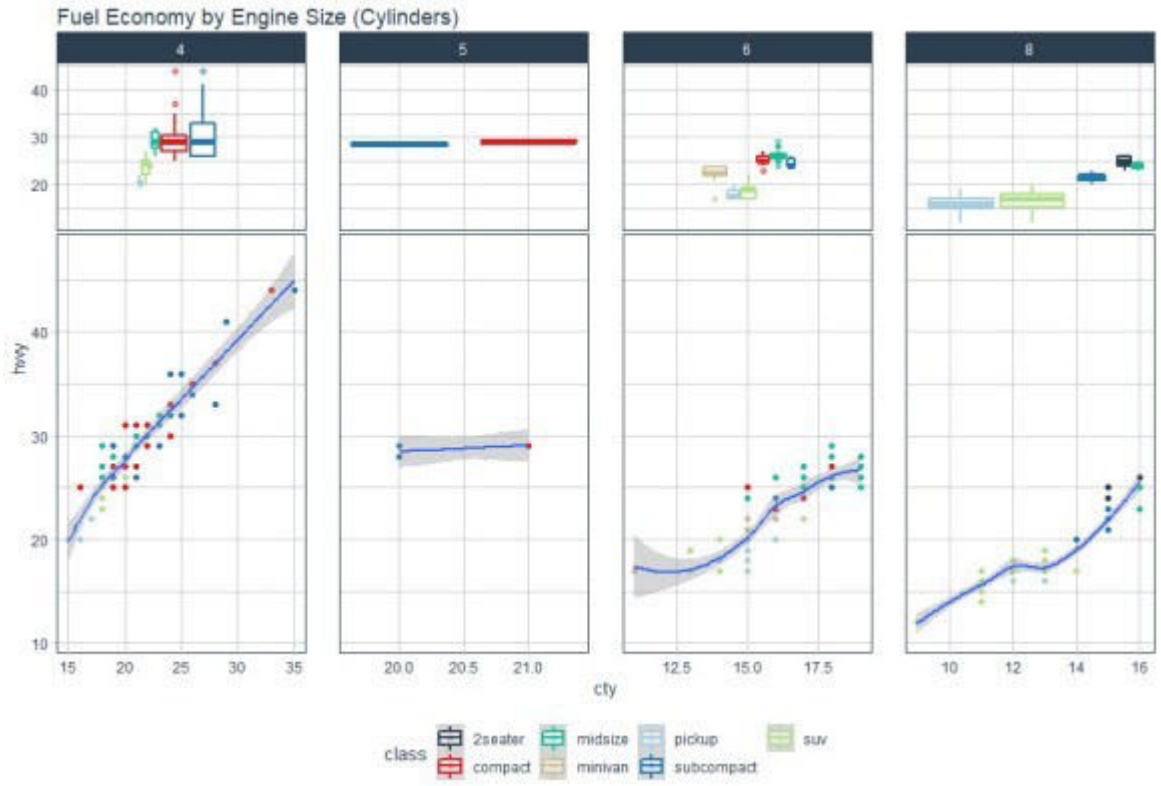
Example 1:

Linear Regression with Marginal Distribution (Density) Side-Plots (Top and Left)



Example 2:

Facet-Plot with Marginal Box Plots (Top)

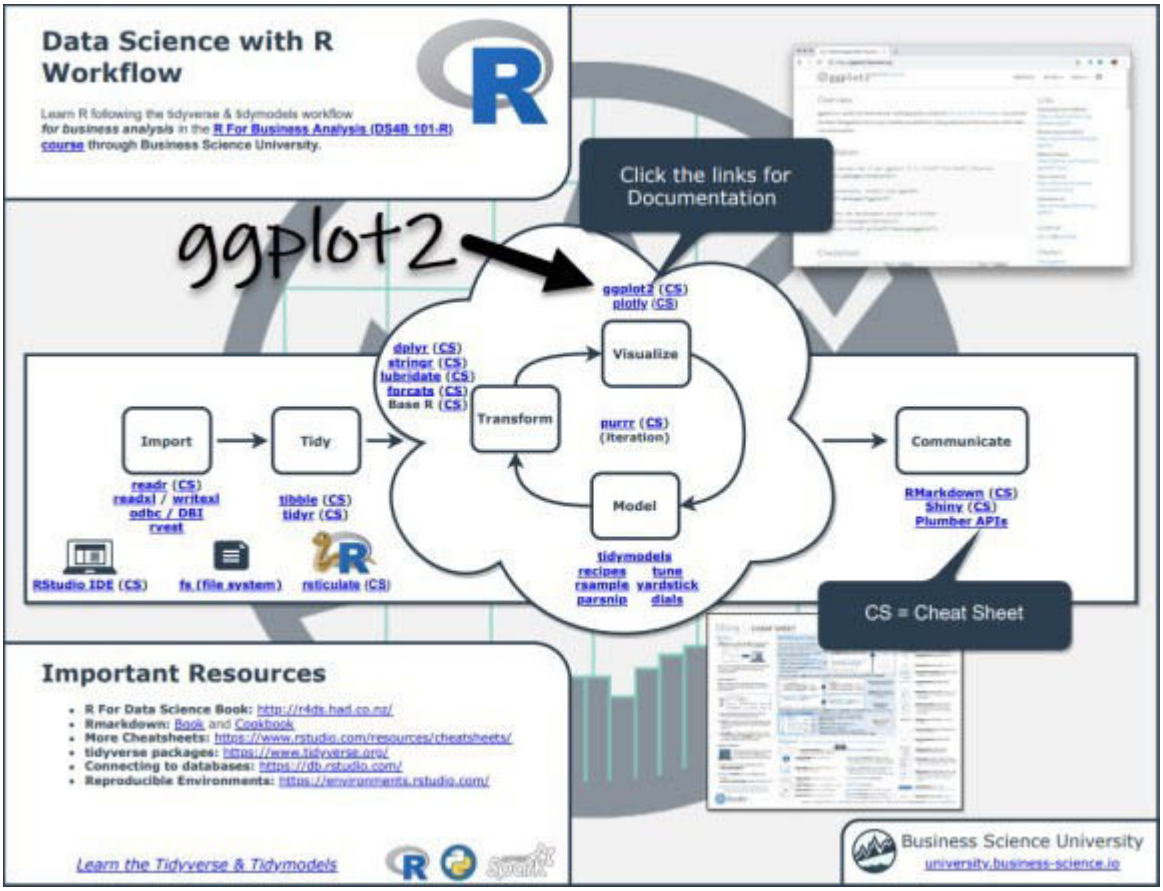


Before we get started, get the Cheat Sheet

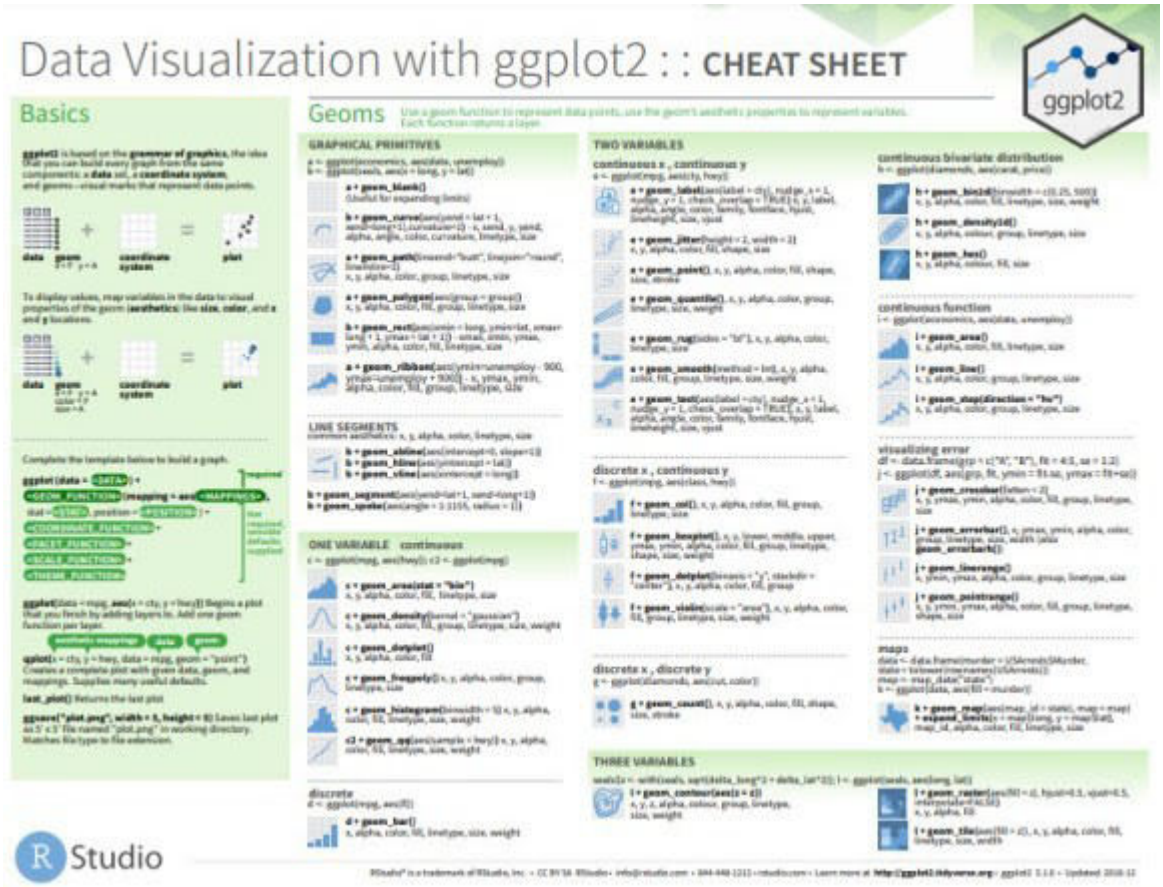
`ggside` is great for making marginal distribution side plots. But, you'll still need to learn how to visualize data with ggplot2. For those topics, I'll use the Ultimate R Cheat Sheet to refer to `ggplot2` code in my workflow.

Quick Example:

Download the Ultimate R Cheat Sheet. **Then Click the "CS" next to "ggplot2"** which opens the Data Visualization with Dplyr Cheat Sheet.



Now you're ready to quickly reference `ggplot2` functions.



Load Libraries & Data

The libraries we'll need today are patchwork, ggridges, ggrepel, maps, tidyverse, and lubridate.

All packages are available on CRAN and can be installed with `install.packages()`. Note – I'm using the development version of `ggside`, which is what I recommend in the YouTube Video .

```
# LIBRARIES ----

# devtools::install_github("jtlandis/ggside")

library(ggside)
library(tidyverse)
library(tidyquant)
```

The dataset is the mpg data that comes with ggplot2.

```
> mpg
# A tibble: 234 x 11
   manufacturer model      displ  year   cyl trans      drv     cty   hwy fl    class
   <chr>        <chr>      <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
 1 audi         a4           1.8  1999     4 auto(15)   f        18    29 p     compact
 2 audi         a4           1.8  1999     4 manual(m5) f        21    29 p     compact
 3 audi         a4           2    2008     4 manual(m6) f        20    31 p     compact
 4 audi         a4           2    2008     4 auto(av)   f        21    30 p     compact
 5 audi         a4           2.8  1999     6 auto(15)   f        16    26 p     compact
 6 audi         a4           2.8  1999     6 manual(m5) f        18    26 p     compact
 7 audi         a4           3.1  2008     6 auto(av)   f        18    27 p     compact
 8 audi         a4 quattro   1.8  1999     4 manual(m5) 4        18    26 p     compact
 9 audi         a4 quattro   1.8  1999     4 auto(l5)   4        16    25 p     compact
10 audi         a4 quattro   2    2008     4 manual(m6) 4        20    28 p     compact
# ... with 224 more rows
```

SLinear Regression with Marginal Distribution Plot

Replicating Seaborn's jointdist() plot

We'll start by replicating what you can do in **Python's Seaborn jointdist() Plot**. We'll accomplish this with `ggside::geom_xsidedensity()`

We set up the plot just like a normal ggplot.

Refer to the Ultimate R Cheat Sheet for:

- `ggplot()`
- `geom_point()`
- `geom_smooth()`

Next we add from ggside:

- `geom_xsidedensity()` – Adds a side density panel (top panel).
- `geom_ysidedensity()` – Adds a side density panel (right panel).

The trick is using the `after_stat(density)`, which makes an awesome looking marginal density side panel plot. I increased the size of the marginal density panels with the `theme(ggside.panel.scale.x)`.

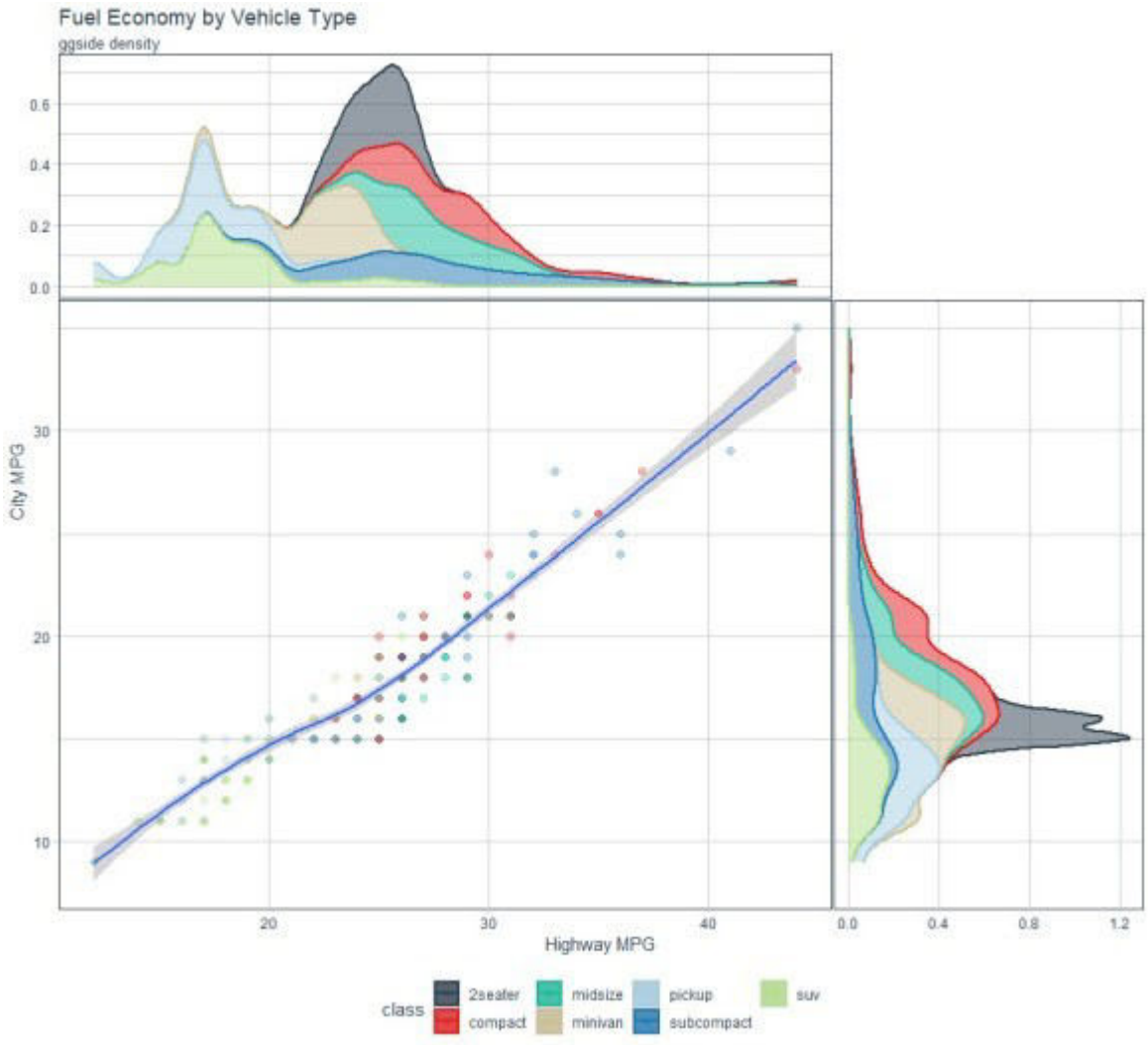
```r
mpg %>%
    ggplot(aes(hwy, cty, color = class)) +
    geom_point(size = 2, alpha = 0.3) +
    geom_smooth(aes(color = NULL), se=TRUE) +
    geom_xsidedensity(
        aes(
            y    = after_stat(density),
            fill = class
        ),
        alpha    = 0.5,
        size     = 1
        ,
        position = "stack"
    ) +
    geom_ysidedensity(
        aes(
            x    = after_stat(density),
            fill = class
        ),
        alpha    = 0.5,
        size     = 1
        ,
        position = "stack"
    ) +
    scale_color_tq() +
    scale_fill_tq() +
    theme_tq() +
    labs(title = "Fuel Economy by Vehicle Type" ,
         subtitle = "ggside density",
         x = "Highway MPG", y = "City MPG") +
    theme(
        ggside.panel.scale.x = 0.4,
        ggside.panel.scale.y = 0.4
    )
```

Get the Code

Loess Regression w/ Marginal Density

We generate the regression plot with marginal distributions (density) to highlight key differences between the automobile classes. We can see:

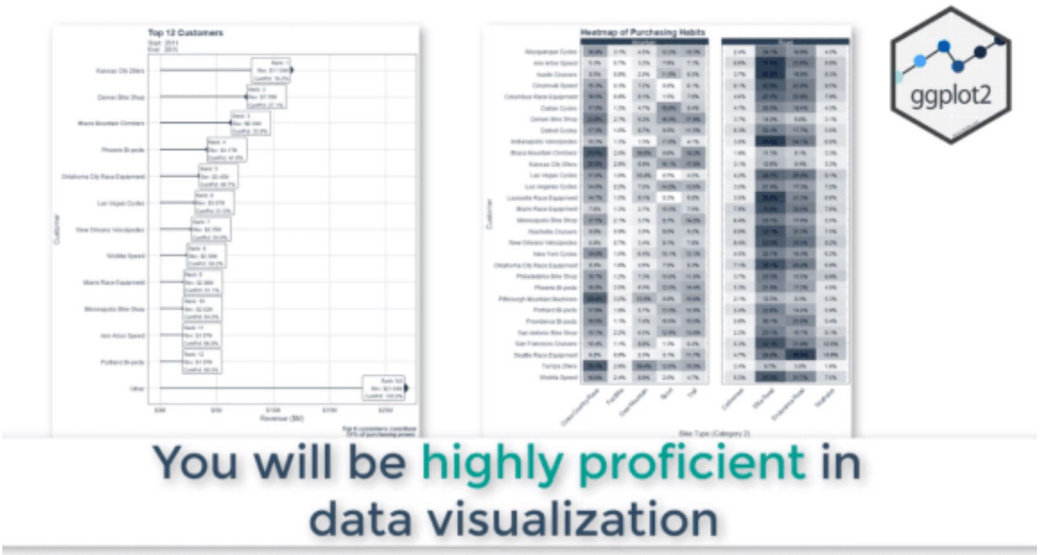
- Pickup, SUV – Have the **lowest** Highway Fuel Economy (MPG)
- 2seater, Compact, Midsize, Subcompact – Have the **highest** Highway Fuel Economy



Need help learning ggplot2?

In the R for Business Analysis (DS4B 101-R) Course , I teach 5-hours just on ggplot2. Learn:

- Geometries
- Scales
- Themes
- And advanced customizations: Labeled Heat Maps and Lollipop Charts



Get started today

Plot 2. Faceted Side-Panels

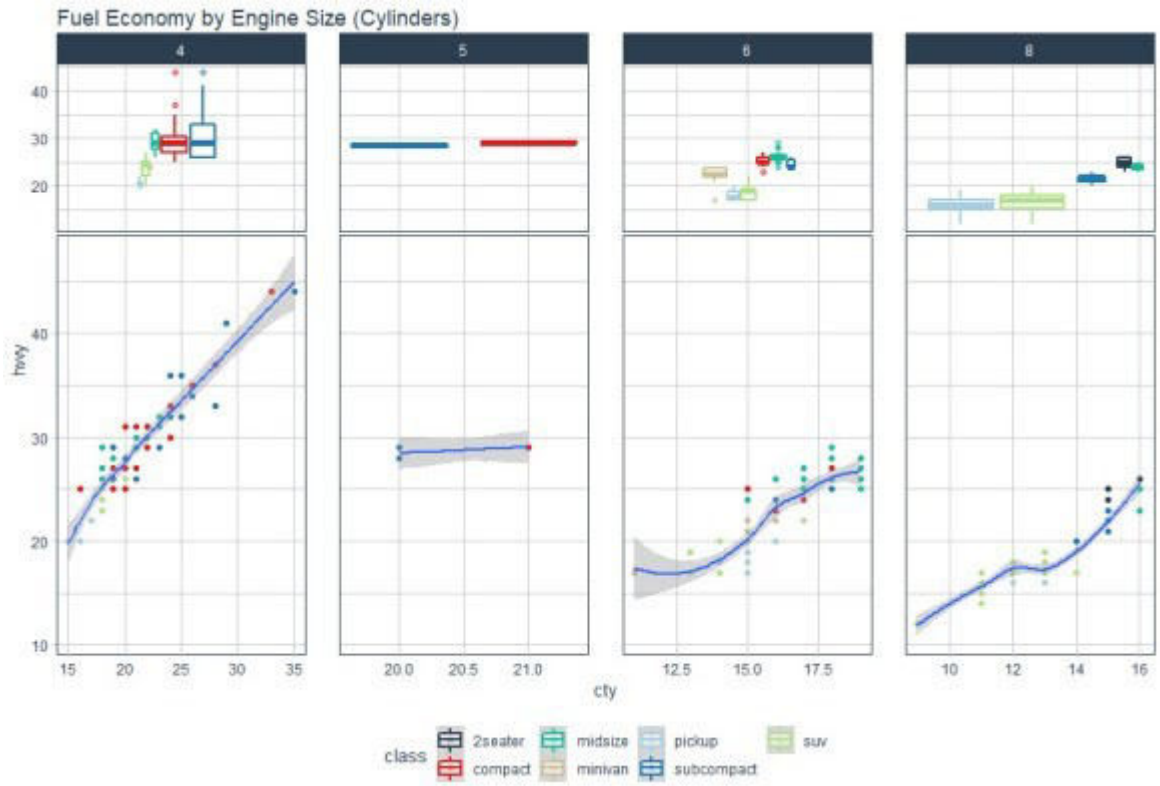
Next, let's try out some advanced functionality. I want to see how ggside handles faceted plots, which are subplots that vary based on a categorical feature. We'll use the "cyl" column to facet, which is for engine size (number of cylinders).

```
mpg %>%
    ggplot(aes(x = cty, y = hwy, color = class)) +
    geom_point() +
    geom_smooth(aes(color = NULL)) +
    geom_xsideboxplot(
        alpha     = 0.5,
        size      = 1
    ) +
    scale_color_tq() +
    scale_fill_tq() +
    theme_tq() +
    facet_grid(cols = vars(cyl), scales = "free_x") +
    labs(
        title = "Fuel Economy by Engine Size (Cylinders)"
    ) +
    theme(
        ggside.panel.scale.x = 0.4
    )
```

Get the Code

Faceted Side Panels? No problem.

Awesome! I have included facets by "cyl", which creates four plots based on the engine size. ggside picked up on the facets and has made 4 side-panel plots.



Amazing. ggside just works.

Congrats. You just quickly made two report-quality plots with ggplot2 and ggside. Excellent work.