

Information about the **tidyverse**

John Hinic

Contents

1	Code to Create this Document	1
2	R Packages for data science	1
3	Some Core Packages	1
3.1	dplyr	2
3.2	ggplot2	2
3.3	readr	3
3.4	tidyr	3

1 Code to Create this Document

```
rmarkdown::render("Hinic_HW2_pdf.Rmd",  
  output_format = "pdf_document",  
  output_options = list(  
    toc = TRUE,  
    toc_depth = "2",  
    number_sections = TRUE,  
    df_print = "tibble"  
  )  
)
```

2 R Packages for data science

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

```
install.packages("tidyverse")
```

3 Some Core Packages

The four *core* packages that we'll use the most are given below along with their purpose and a quick example of some functionality.

3.1 dplyr



[dplyr is a grammar of data manipulation](#), providing a consistent set of verbs that help you solve the most common data manipulation challenges:

- `mutate()` adds new variables that are functions of existing variables.
- `select()` picks variables based on their names.
- `filter()` picks cases based on their values.
- `summarise()` reduces multiple values down to a single summary
- `arrange()` changes the ordering of the rows

These all combine naturally with `group_by()` which allows you to perform any operation “by group”. You can learn more about them in `vignette(“dplyr”)`. As well as these single-table verbs, dplyr also provides a variety of two-table verbs, which you can learn in `vignette(“two-table”)`.

If you are new to dplyr, the best place to start is the data transformation chapter in *R for data science*.

```
library(dplyr)

starwars %>%
  filter(species == "Droid")
```

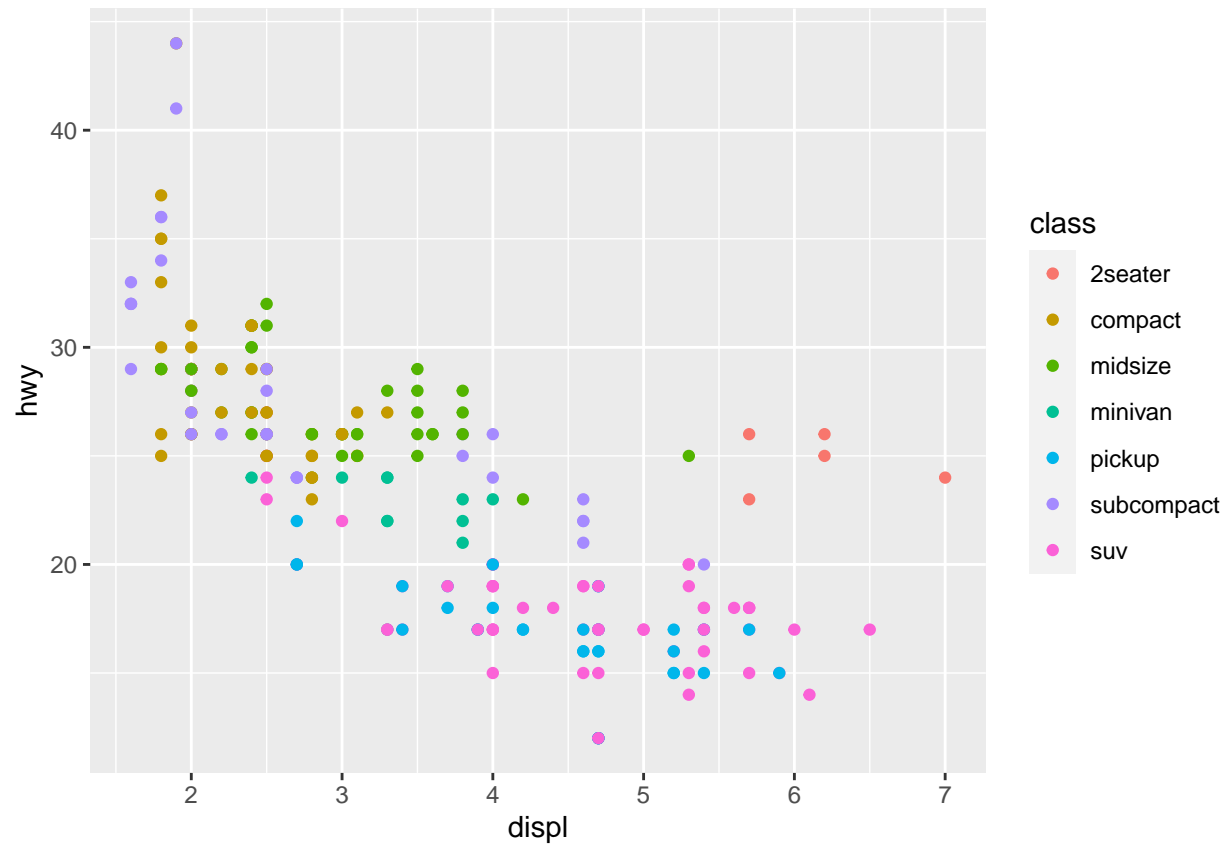
3.2 ggplot2



[ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics](#). You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

```
library(ggplot2)

ggplot(mpg, aes(displ, hwy, colour = class)) +
  geom_point()
```

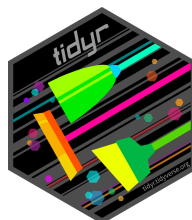


3.3 readr



The goal of [readr](#) is to provide a fast and friendly way to read rectangular data (like csv, tsv, and fwf). It is designed to flexibly parse many types of data found in the wild, while still cleanly failing when data unexpectedly changes. If you are new to readr, the best place to start is the data import chapter in R for data science.

3.4 tidyr



The goal of `tidyr` is to help you create tidy data. Tidy data is data where:

1. Every column is variable.
2. Every row is an observation.
3. Every cell is a single value.

Tidy data describes a standard way of storing data that is used wherever possible throughout the tidyverse. If you ensure that your data is tidy, you'll spend less time fighting with the tools and more time working on your analysis. Learn more about tidy data in `vignette("tidy-data")`.

```
library(tidyr)
```

```
relig_income
```

```
relig_income %>%  
  pivot_longer(-religion, names_to = "income", values_to = "frequency")
```