Portfolio 5

Sam Bowyer

The Tidyverse

On tidyverse.org, the tidyverse is described as "an opinionated collection of R packages designed for data science" which "share an underlying deign philosophy, grammar and data structures". This collection of packages is quite large (see tidyverse.org/packages), however, the main eight packages are:

- `ggplot2` - for creating graphics
- `dplyr` - for data manipulation
- `tidyr` - for tidying data
- `readr` - for reading data files (such as csv, tsv etc.)
- `purrr` - for extending `R`'s functional programming capabilities
- `tibble` - a useful implementation of dataframes
- `stringr` - for working with strings
- `forcats` - for working with factors (categorical data)

We'll go through an example of working with some of these packages that will give an idea of how the tidyverse is used for data analysis, however, first it will be useful to work with another package: `magrittr`.
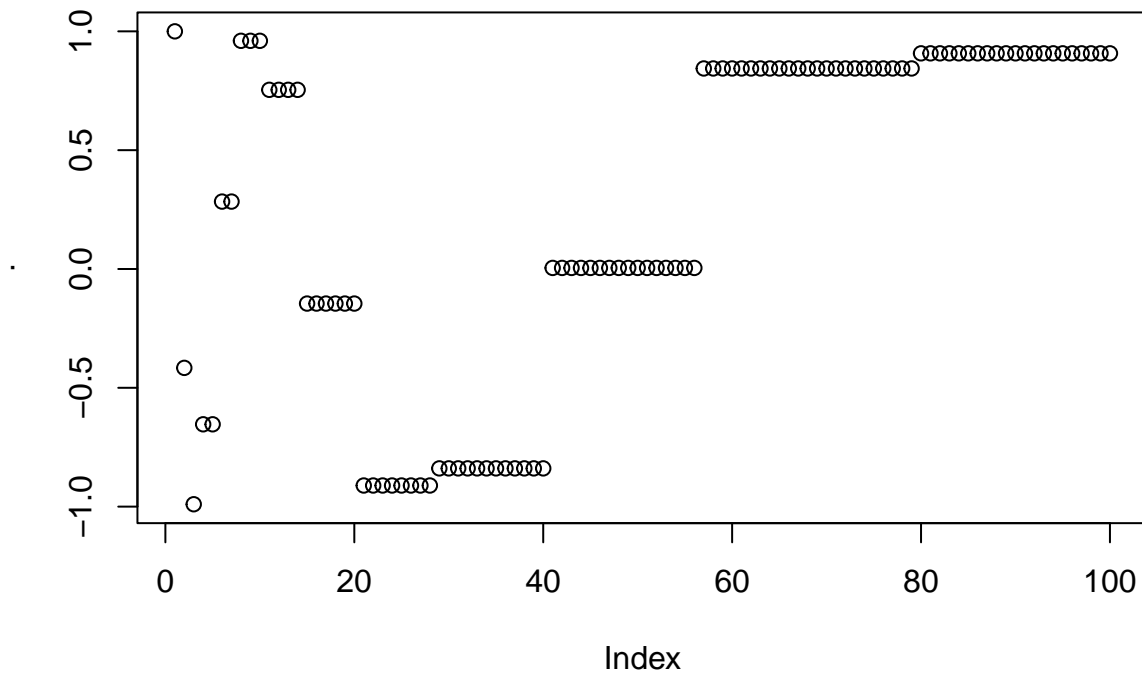
Pipes in `magrittr`

The package `magrittr` allows us to use the pipe operator `%>%` which can simplify and make code more readable, for example we can rewrite this code

```
x = 1:100
plot(cos(floor(log(x, 1.4))))
```
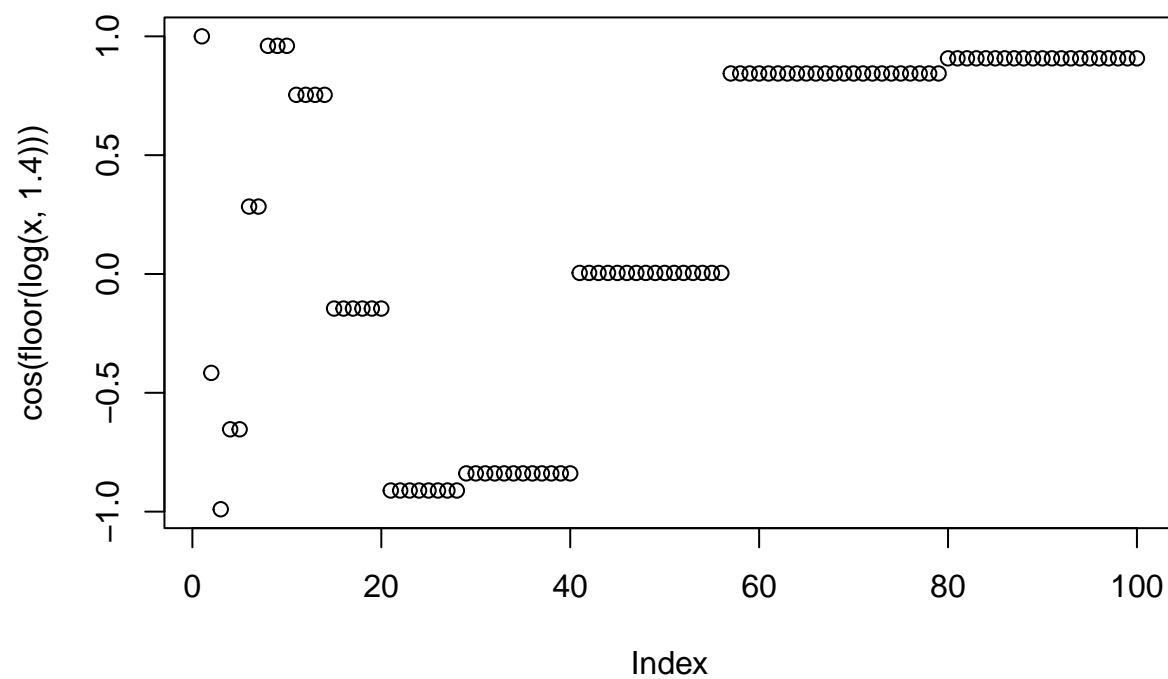
as the following

```
library(magrittr)

x %>% log(. , 1.4) %>% floor %>% cos %>% plot
```

So the output of whatever is to the left of the pipe gets sent as input to the expression on the right of the pipe, and as we saw with the `log(., 1.4)` part of the second example we can reference the left side's output directly with "`.`" if we want to specify other arguments to a function that we're using.

Data Analysis With The Tidyverse

example from lab

Pivoting

One further example we'll quickly consider is using `dplry` and `tidyr` to reshape some data. In particular we'll use the