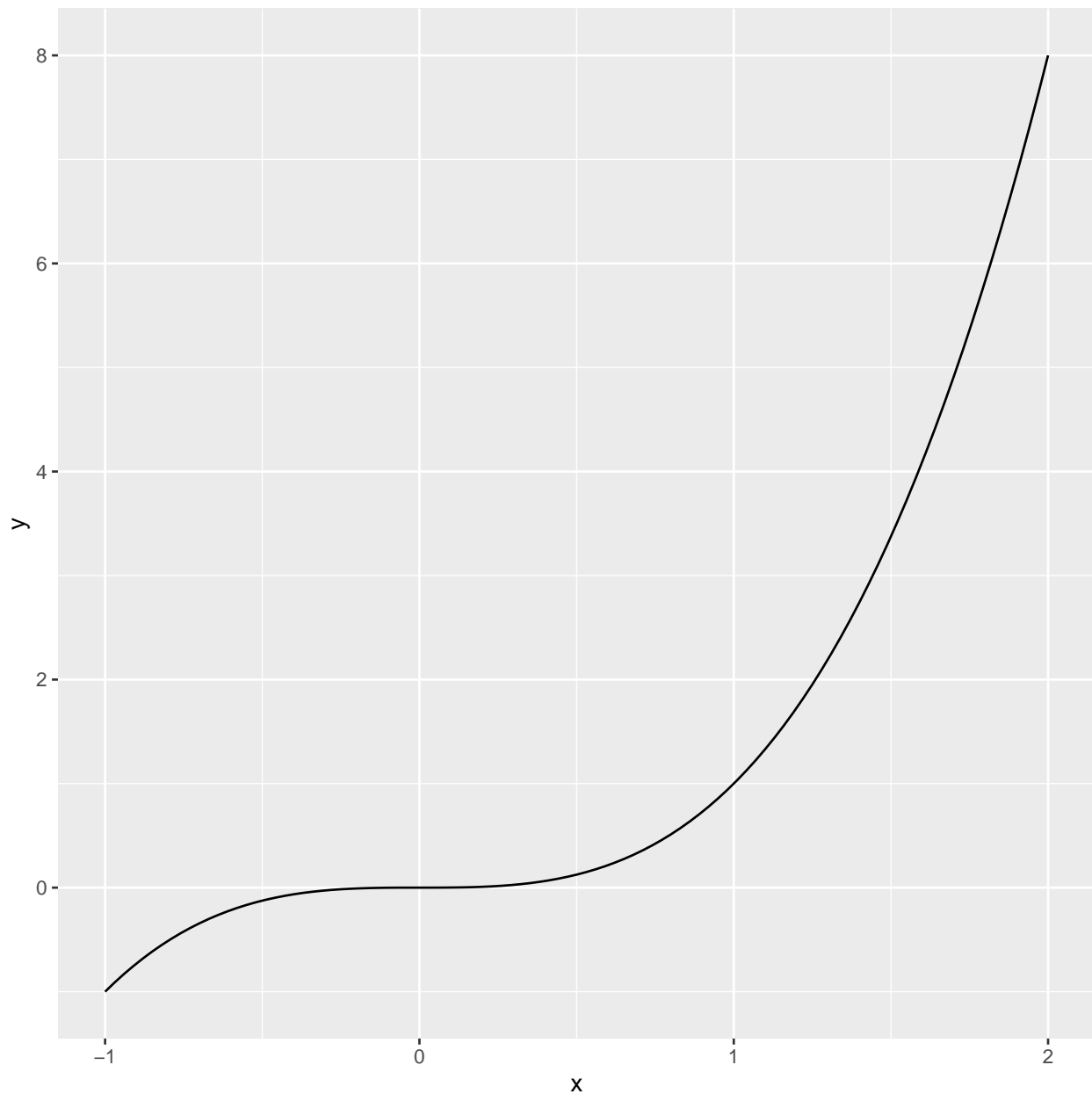


Intro to Data Science
Homework 3: Due Wednesday September 25 at 2:00pm

Exercises:

1. Go through section 9 Functions of the R Programming course in the `swirl` package, then answer the following questions:
 - (a) What is a function?
 - (b) What does the `Sys.Data()` function do? How many input arguments are required?
 - (c) What are the two “slogans” for R stated by John Chambers?
 - (d) How do you see the source code for an R function?
 - (e) Why would having default arguments be useful?
 - (f) What does the `args` function do? Give an example of its use.
 - (g) Explain why one might want to pass a function as an argument to another function.
 - (h) What is an easy way to return the last element of an arbitrary vector?
 - (i) What does the `paste` function do?
 - (j) What is the significance of the “dot-dot-dot” argument for a function in R?
2. Write an R function that inputs a vector and computes the mean of the vector. Save your function in an R script called `my_mean_func.R`. Be sure to test your function and make sure it is working correctly.
3. Write an R function that inputs two whole numbers and returns the remainder after dividing the first by the second. Save your function in an R script called `my_remain_func.R`. Be sure to test your function and make sure it is working correctly.
4. This exercise walks you through the steps to plot the graph of a function $y = f(x)$ using the `stat_function()` function from the `ggplot2` package. Here are the commands to plot $y = x^3$ over the interval $[-1, 2]$:

```
library(ggplot2)
# create function for y=x^3
y_cubed <- function(x){
  x^3
}
# create input values
x <- seq(-1,2,by=0.05)
f_df <- data.frame(x=x)
# plot graph
ggplot(f_df, aes(x=x)) + stat_function(fun="y_cubed")
```



(a) Using the previous code as a template, plot the specified functions over the given interval:

- i. $y = 2x^2 - 3x + 5$, $[-2, 6]$
- ii. $y = e^{-2x}$, $[0, 3]$
- iii. $y = \ln(x)$, $[0.5, 3]$

5. What trigonometric functions does R provide? Make plots of each of the trigonometric functions over an appropriate period.
6. Using the `flights` data from the `nycflights13` package, find all flights that
- (a) Had an arrival delay of two or more hours
 - (b) Flew to Houston (IAH or HOU)
 - (c) Were operated by United, American, or Delta

- (d) Departed in summer (July, August, and September)
 - (e) Arrived more than two hours late, but didn't leave late
 - (f) Were delayed by at least an hour, but made up over 30 minutes in flight
 - (g) Departed between midnight and 6am (inclusive)
7. Another useful **dplyr** filtering helper is **between()**. What does it do? Can you use it to simplify the code needed to answer the previous challenges?
 8. How many flights have a missing `dep_time`? What other variables are missing? What might these rows represent?
 9. How could you use `arrange()` to sort all missing values to the start? (Hint: use `is.na()`).
 10. Sort `flights` to find the most delayed flights. Find the flights that left earliest.
 11. Sort `flights` to find the fastest flights.