

# GENOME 569A

Class 5: Electronic lab notebooks

# Project 1 questions

# Why make electronic lab notebooks?

In a perfect world, you'd have a single pipeline that, when run, generates all figures, tables, and statistics for your paper

An electronic lab notebook makes your analysis **transparent**, **unambiguous**, and **replicable** to a reader

Automating analysis can **save time** in the long run

Allows you to **version control** your science



# Markdown

# Problem: Make a pretty website, but only spend time on content

## Solution: Markdown

### ## Course description

Programming skills and software tools for building automated bioinformatics pipelines and computational biology analyses. Emphasis on UNIX tools and R libraries for distilling raw sequencing data into interpretable results. This course is aimed at students familiar with UNIX and with some programming experience in python, R, or C/C++.

### ## Instructional staff

- \* [Cole Trapnell](http://cole-trapnell-lab.github.io/)
- \* Teaching assistant: April Lo

Please click on the links above for email addresses and office locations.

### ## Meeting times and locations

#### ### Classes:

Monday and Wednesday, 10:30 – 12:20 am, [zoom](http://www.washington.edu/home/maps/southcentral.html?gnom).

#### ### Class Slack:

We will use Slack during class and outside of class to communicate, share code snippets, ask and answer questions. The class slack is here:

- \* [gs-bioinf-workflows.slack.com](https://gs-bioinf-workflows.slack.com)

You will receive an invitation to join prior to the first class.

#### ### Office hours:

- \* No official office hours. Post questions on Slack as needed.

### ## Prerequisites

- \* Substantial background in molecular and cellular biology, genetics, biochemistry, or related disciplines.
- \* Familiarity with UNIX.
- \* Some programming experience in python, R, or C/C++.
- \* Students are encouraged to have taken GENOME559 and/or GENOME560.

# genome569

[View on GitHub](#)

### Course description

Programming skills and software tools for building automated bioinformatics pipelines and computational biology analyses. Emphasis on UNIX tools and R libraries for distilling raw sequencing data into interpretable results. This course is aimed at students familiar with UNIX and with some programming experience in python, R, or C/C++.

### Instructional staff

- Cole Trapnell
- Teaching assistant: April Lo

Please click on the links above for email addresses and office locations.

### Meeting times and locations

#### Classes:

Monday and Wednesday, 10:30 - 12:20 am, [zoom](#).

#### Class Slack:

We will use Slack during class and outside of class to communicate, share code snippets, ask and answer questions. The class slack is here:

- [gs-bioinf-workflows.slack.com](#)

You will receive an invitation to join prior to the first class.

# Markdown file structure

```
---  
title: "genome569"  
output: html_document  
---
```

## **## Course description**

Programming skills and software tools for building automated bioinformatics pipelines and computational biology analyses. Emphasis on UNIX tools and R libraries for distilling raw sequencing data into interpretable results. This course is aimed at students familiar with UNIX and with some programming experience in python, R, or C/C++.

## **## Instructional staff**

- \* [Cole Trapnell](<http://cole-trapnell-lab.github.io/>)
- \* Teaching assistant: April Lo



# Markdown file structure

```
---  
title: "genome569"  
output: html_document  
---
```

“Front matter”: Meta data  
about the document goes  
here

## ## Course description

Programming skills and software tools for building automated bioinformatics pipelines and computational biology analyses. Emphasis on UNIX tools and R libraries for distilling raw sequencing data into interpretable results. This course is aimed at students familiar with UNIX and with some programming experience in python, R, or C/C++.

## ## Instructional staff

- \* [Cole Trapnell](<http://cole-trapnell-lab.github.io/>)
- \* Teaching assistant: April Lo

# Markdown file structure

```
---  
title: "genome569"  
output: html_document  
---
```

**## Course description** ← Section header

Programming skills and software tools for building automated bioinformatics pipelines and computational biology analyses. Emphasis on UNIX tools and R libraries for distilling raw sequencing data into interpretable results. This course is aimed at students familiar with UNIX and with some programming experience in python, R, or C/C++.

← Paragraph

**## Instructional staff**

\* [Cole Trapnell](<http://cole-trapnell-lab.github.io/>)

\* Teaching assistant: April Lo

← Hyperlink



# Markdown file structure

```
---  
title: "genome569"  
output: html_document  
---
```

## ## Course description

Programming skills and software tools for building automated bioinformatics pipelines and computational biology analyses. Emphasis on UNIX tools and R libraries for distilling raw sequencing data into interpretable results. This course is aimed at students familiar with UNIX and with some programming experience in python, R, or C/C++.

## ## Instructional staff

- \* [Cole Trapnell](<http://cole-trapnell-lab.github.io/>)
- \* Teaching assistant: April Lo

**Bulleted list. Enumerated lists also easy.**

# Markdown file structure

```
---  
title: "genome569"  
output: html_document  
---
```

## ## Course description

Programming skills and software tools for building automated bioinformatics pipelines and computational biology analyses. Emphasis on UNIX tools and R libraries for distilling raw sequencing data into interpretable results. This course is aimed at students familiar with UNIX and with some programming experience in python, R, or C/C++.

## ## Instructional staff

- \* [Cole Trapnell](<http://cole-trapnell-lab.github.io/>)
- \* Teaching assistant: April Lo

**Bulleted list. Enumerated lists also easy.**

# Markdown key features

Markdown is a lightweight, plain text markup language that can be rendered into any number of formats

Allows you to forget about formatting and focus on content

Rendering engines for HTML, PDF, MS word, many others are available.

Some web hosting and blogging services (including GitHub) support posts in Markdown

Can run code written in other language and include results inline to generate data-driven reports



# Rmarkdown

# Problem: Make a data-driven report (or electronic lab notebook)

## Solution: Rmarkdown

```
---
title: "test2"
output: html_document
---
```

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
```

### ## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
```{r cars}
summary(cars)
```
```

### ## Including Plots

You can also embed plots, for example:

```
```{r pressure, echo=FALSE}
plot(pressure)
```
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

## test2

### R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

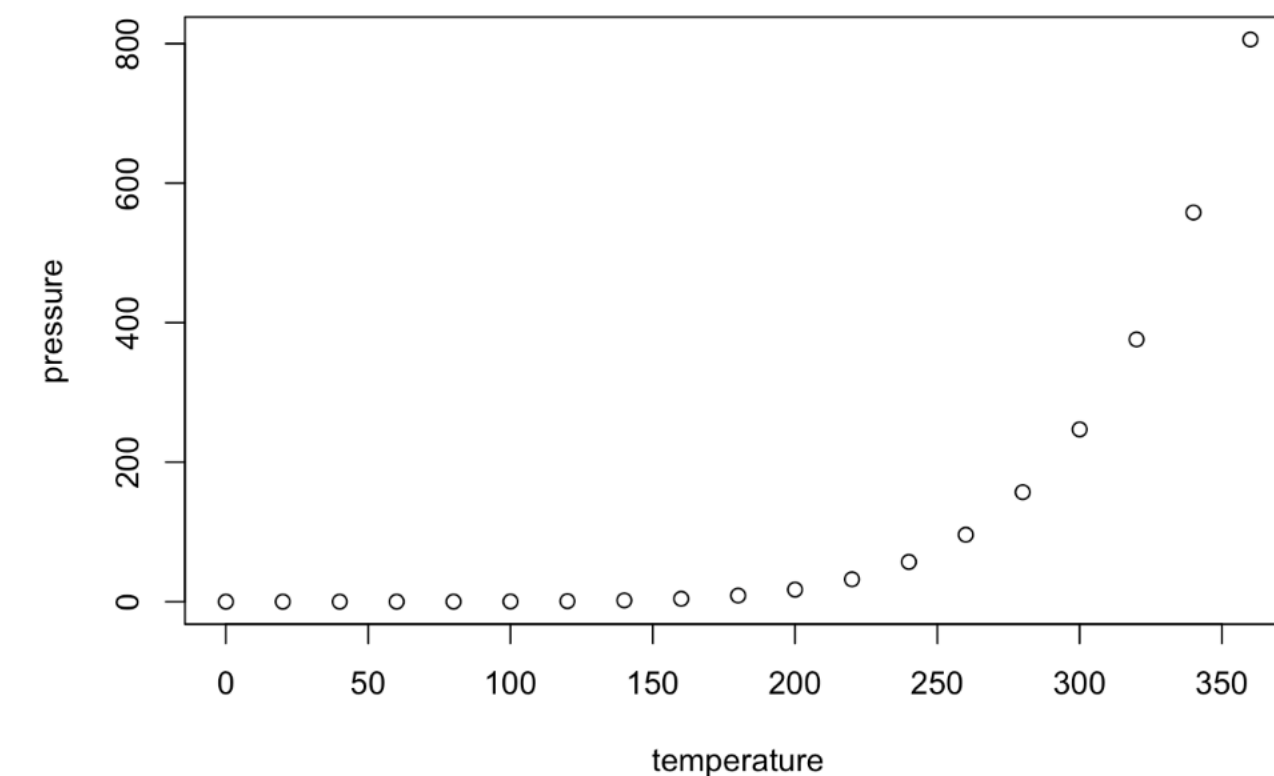
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

### Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

# Rmarkdown embeds R code in Markdown

```
```{r cars}  
summary(cars)  
```
```

```
summary(cars)
```

```
##      speed      dist  
##  Min.   : 4.0    Min.   : 2.00  
## 1st Qu.:12.0    1st Qu.: 26.00  
## Median :15.0    Median : 36.00  
## Mean   :15.4    Mean   : 42.98  
## 3rd Qu.:19.0    3rd Qu.: 56.00  
## Max.   :25.0    Max.   :120.00
```



# How Rmarkdown works



Image credit: [www.rstudio.com](http://www.rstudio.com)

# Rstudio has a slick notebook interface for Rmarkdown





# Problem 1

Create a new R notebook in Rstudio (File -> New File -> R Notebook)

Follow the instructions in the template to run and add code chunks

In a new chunk, compute the average speed and distance for the cars data frame and print the results

Render the Notebook to HTML, PDF, and MS Word