OLE Final Project

2024-08-21

Project: Transcription factor target gene prediction through multi-omics datasets

Aim: Prediction is a central application in many real-world data analyses. In this project, we will aim to apply classification techniques for predicting novel transcription factor target genes.

Background: Transcriptional regulation is a vital process in all living organisms. It is orchestrated by transcription factors along with other factors that in concert drive mRNA expression. Distinct transcriptional networks drive development, lineage specification, and cell fate decisions during early embryonic development (Theunissen and Jaenisch, 2017). Recent advances in omics technologies have made it possible to profile genome-wide transcriptional and epigenetic events for investigating transcriptional networks. A key goal is to identify the target genes of transcription factors that drive key transcriptional events over a course of time.

```
# Load necessary libraries
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(randomForest)
```

```
## randomForest 4.7-1.1
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
##
## The following object is masked from 'package:dplyr':
##
##     combine
##
## The following object is masked from 'package:ggplot2':
```

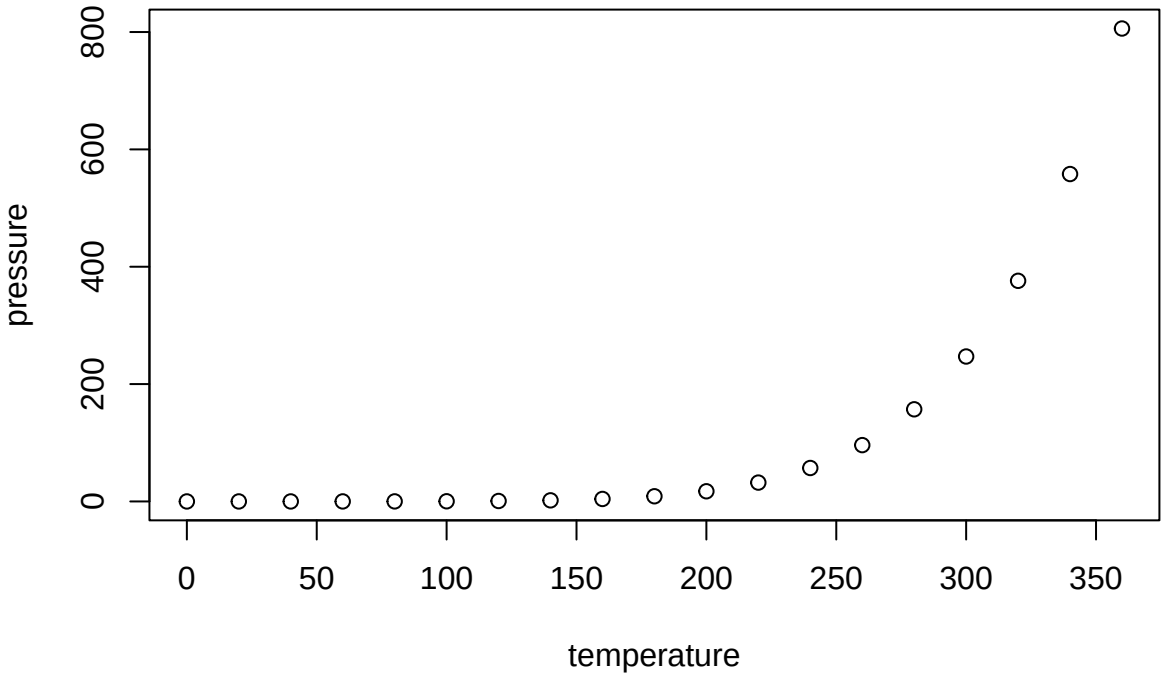
```
##
##     margin
```

```
library(e1071)
library(ggplot2)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

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.