High Dimensional Statistics IV MATH4287 - Mini Project Report

qvns53

2025-12-16

R Markdown

Yeung and Ruzzo (2001) Hubert and Arabie (1985) Mukherjee and Zhang (2022) Ding and He (2004) Rand (1971) Chang (1983)

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.

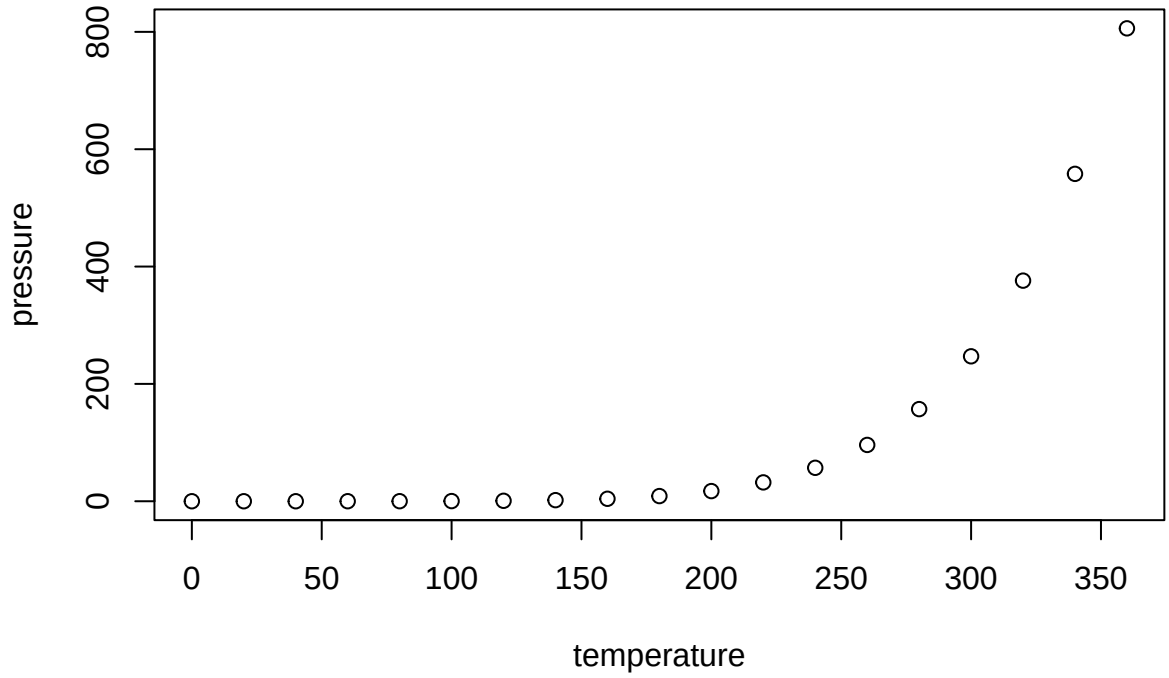
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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.

Chang, Wei-Chien. 1983. "On Using Principal Components Before Separating a Mixture of Two Multivariate Normal Distributions." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 32 (3): 267–75. https://doi.org/10.2307/2347949.

Ding, Chris, and Xiaofeng He. 2004. "K -Means Clustering via Principal Component Analysis." In *Twenty-First International Conference on Machine Learning - ICML '04*, 29. Banff, Alberta, Canada: ACM Press. https://doi.org/10.1145/1015330.1015408.

Hubert, Lawrence, and Phipps Arabie. 1985. "Comparing Partitions." *Journal of Classification* 2 (1): 193–218. https://doi.org/10.1007/BF01908075.

Mukherjee, Chandra, and Jiapeng Zhang. 2022. "Compressibility: Power of PCA in Clustering Problems Beyond Dimensionality Reduction." *ResearchGate*. https://www.researchgate.net/publication/360186068_Compressibility_Power_of_PCA_in_Clustering_Problems_Beyond_Dimensionality_Reduction.

Rand, William M. 1971. "Objective Criteria for the Evaluation of Clustering Methods." *Journal of the American Statistical Association* 66 (336): 846–50. https://doi.org/10.1080/01621459.1971.10482356.

Yeung, Ka Yee, and Walter L Ruzzo. 2001. "An Empirical Study on Principal Component Analysis for Clustering Gene Expression Data."