Introductory Data Science

Assignment 1

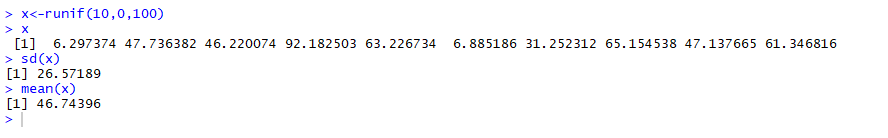
Haitian Yu

Question 1:

* R is a language and environment that is used for both statistical analysis and graphical representation of data. It is widely used by both statisticians, data scientists, and data miners to develop statistical software and data analysis.
* 4 Things You Can Do With R:
  + Linear Modelling
  + Non-Linear Modelling
  + Time-Series Analysis
  + Classical Statistical Tests
* 2 Online Resources to Help Learn R:
  + Datacamp’s Free R Tutorial: <https://www.datacamp.com/onboarding/learn?from=home&technology=r>
  + Stackoverflow Forums: <https://stackoverflow.com/questions/tagged/r>

Question 2:

a)



b)

|  |
| --- |
|  |
| <https://stackoverflow.com/questions/17772505/create-a-100-number-vector-with-random-values-in-r-rounded-to-2-decimals>  <http://www.r-tutor.com/elementary-statistics/numerical-measures/standard-deviation> |
| |  | | --- | |  | |

Question 3:

1. install.packages(“rmarkdown”)
2. R Markdown provides an authoring framework for data science. It can be used to save and execute code and generate high quality reports for sharing.
3. See RMD File
4. See RMD File
5. R Notebooks allows you to keep everything in a single document, whether it’s scripts, output, or even a narrative. After you have built everything, R Notebooks allows you to publish the work as a report.