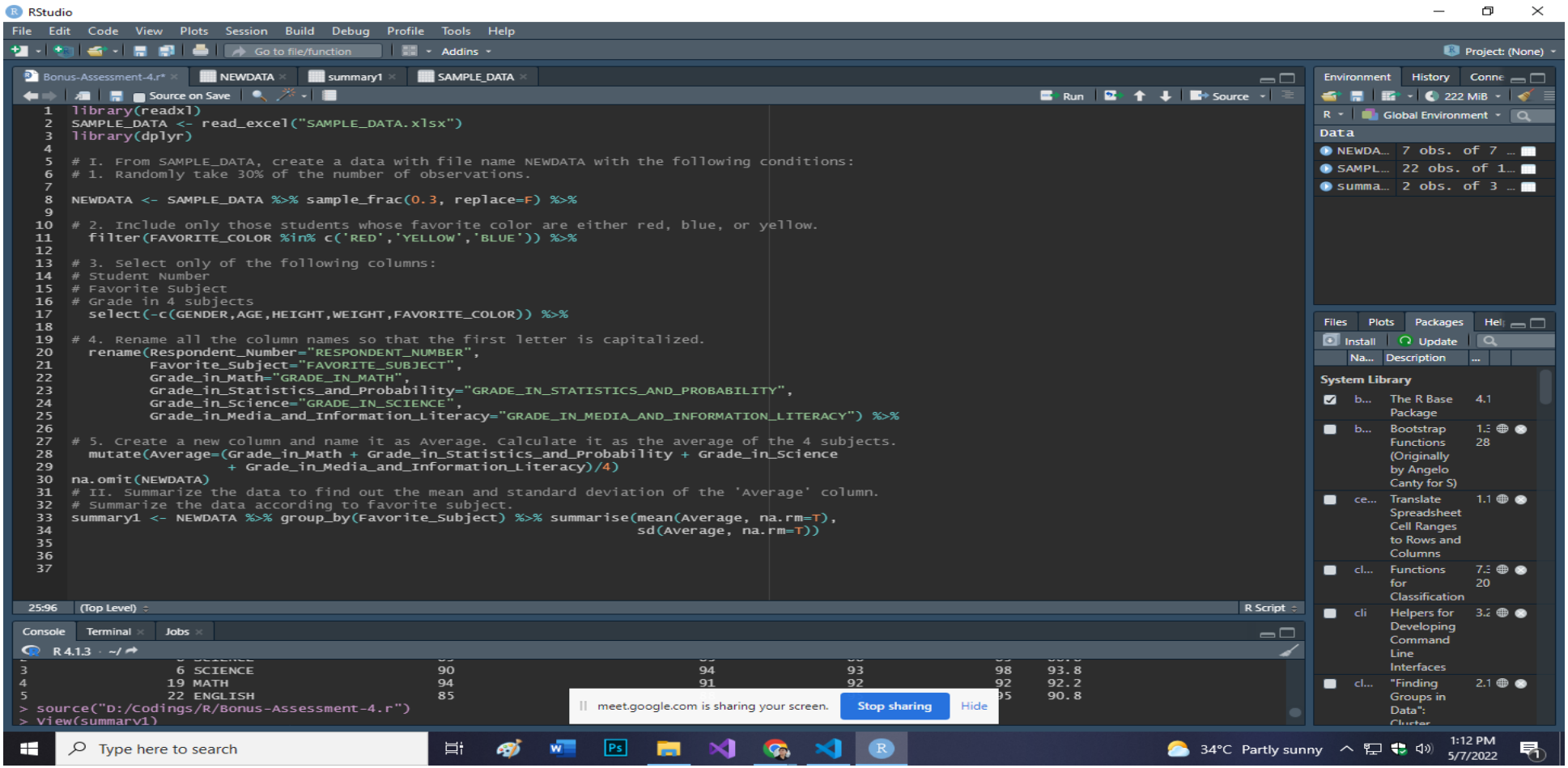
ENGINEERING DATA ANALYSIS DOCUMENTATION

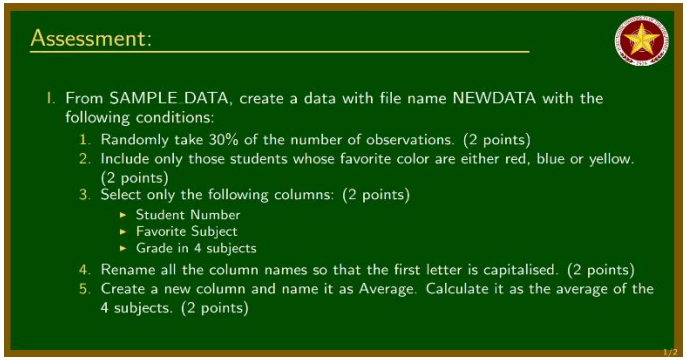
TOPIC: DATA MANIPULATION (BONUS ASSESSMENT 4)	**COURSE AND SECTION: BSCOE 1-1**
PROFESSOR: PERLYN DILLA	**DATE: MAY 7, 2022**

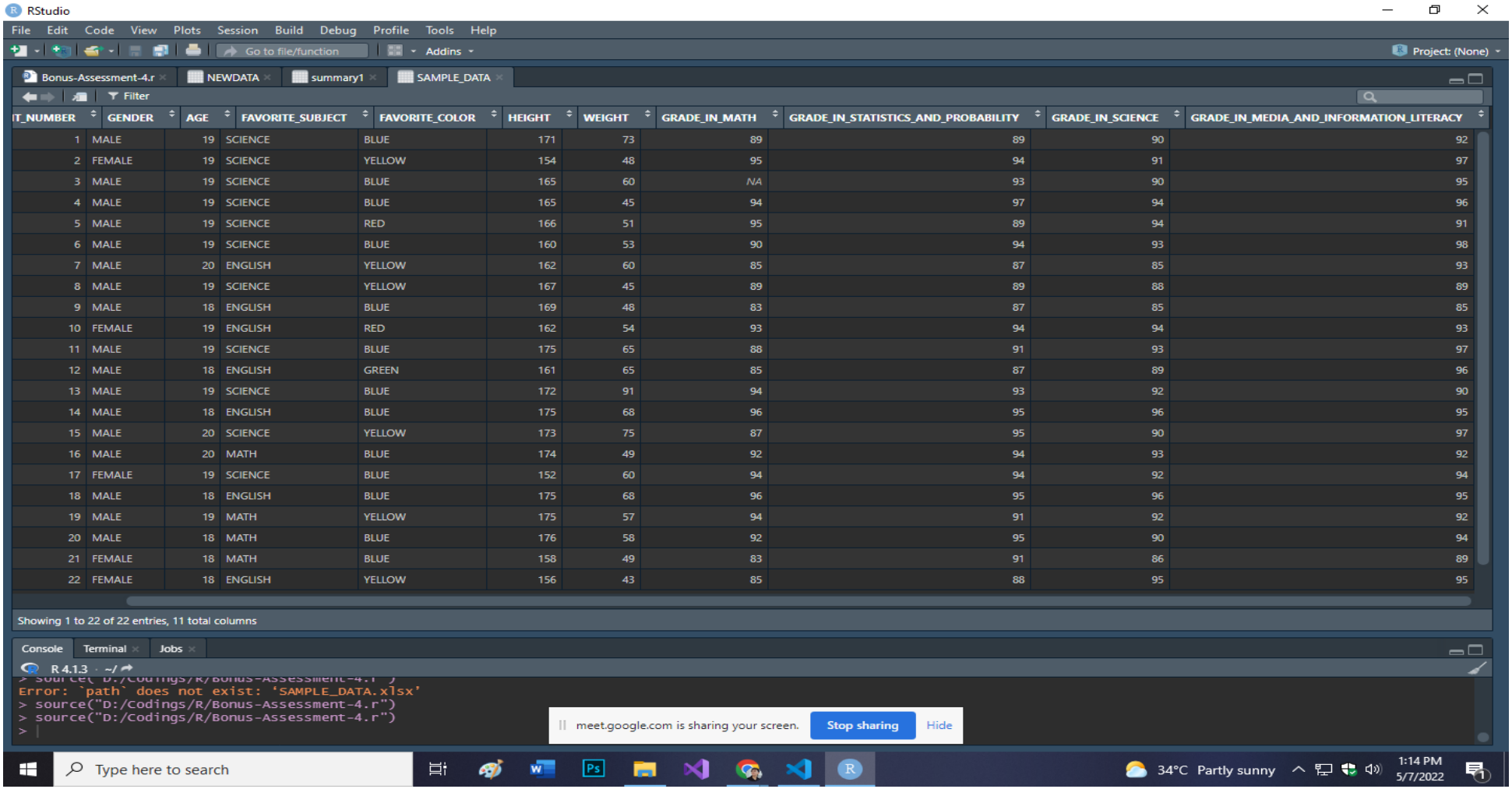
GROUP MEMBERS:

BUHAIN, JASON
CABATE, REIVEN
DELFIN, DANIELLA
RAMOS, ARVELYN
VIERNES, MICHAEL

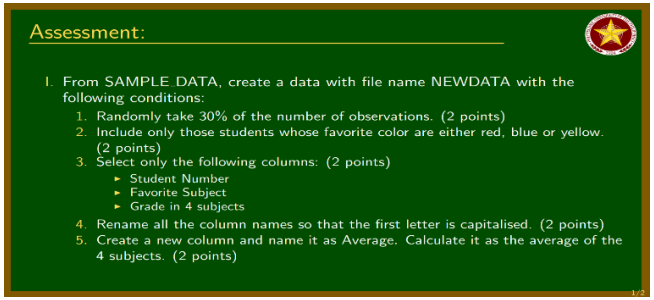


- **SOURCE CODE FOR THE PROGRAMMING EXERCISES**





THE SAMPLE DATA



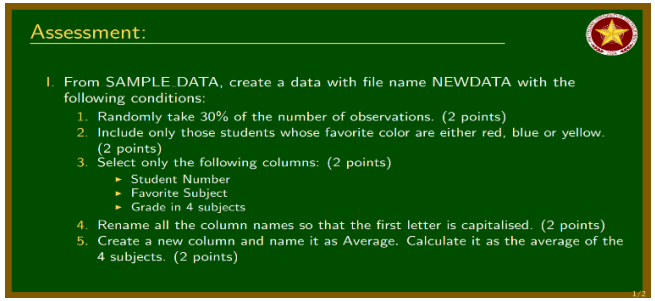
RGui (64-bit) - [Data: NEWDATA1]

File

	RESPONDENT_NUMBER	GENDER	AGE	FAVORITE_SUBJECT	FAVORITE_COLOR	HEIGHT	WEIGHT	GRADE_IN_MATH
1	8	MALE	19	SCIENCE	YELLOW	167	45	89
2	2	FEMALE	19	SCIENCE	YELLOW	154	48	95
3	18	MALE	18	ENGLISH	BLUE	175	68	96
4	16	MALE	20	MATH	BLUE	174	49	92
5	10	FEMALE	19	ENGLISH	RED	162	54	93
6	4	MALE	19	SCIENCE	BLUE	165	45	94
7	20	MALE	18	MATH	BLUE	176	58	92

9:56 PM

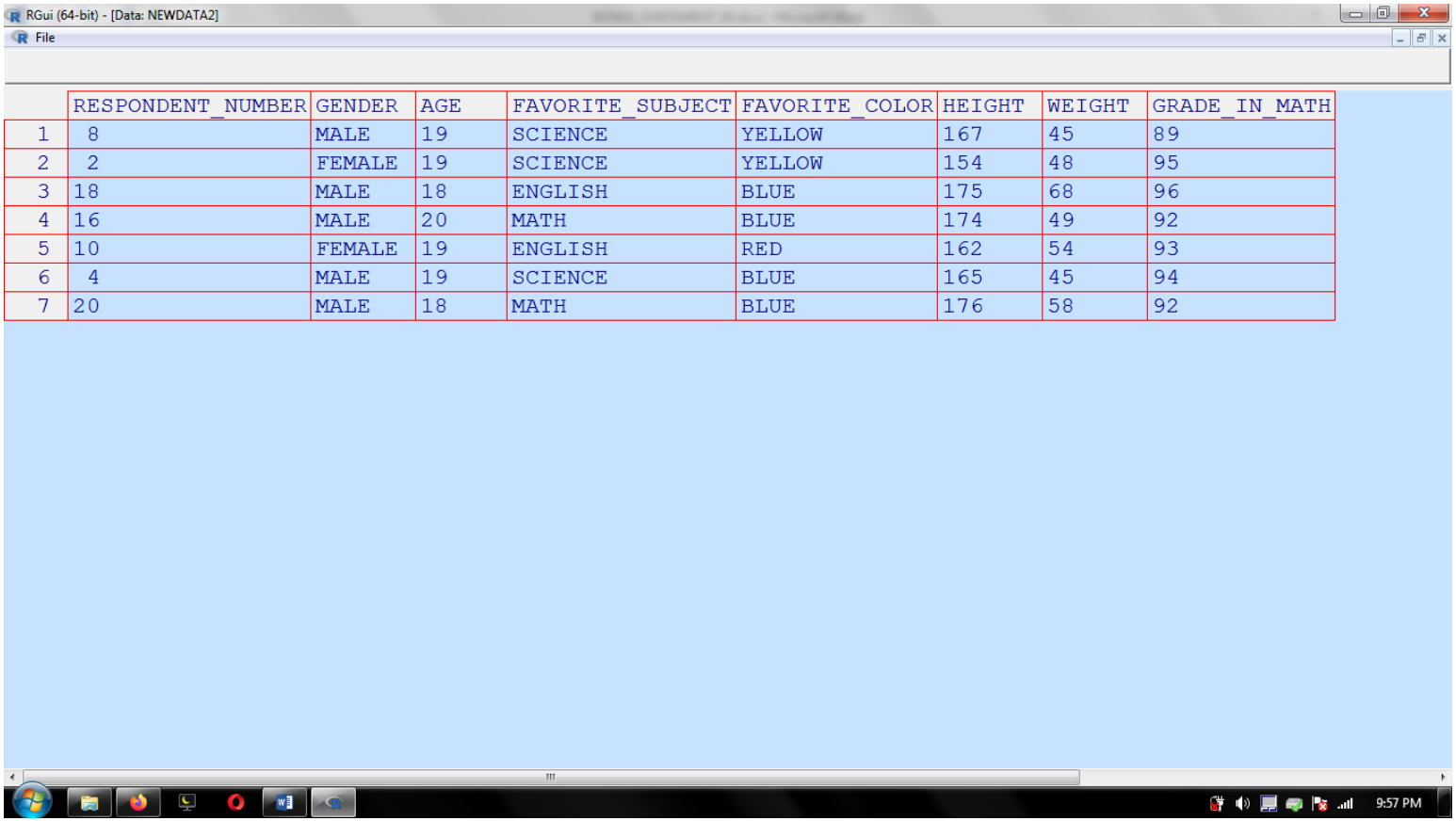
1. Randomly take 30% of the number of observations.



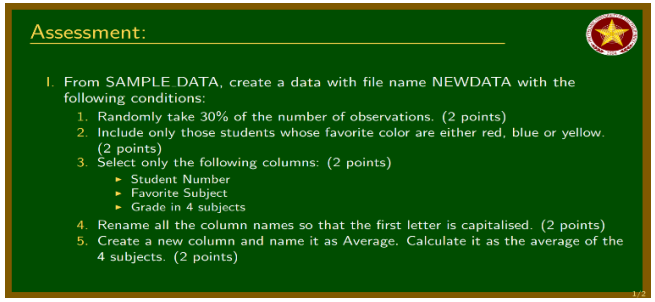
RGui (64-bit) - [Data: NEWDATA2]

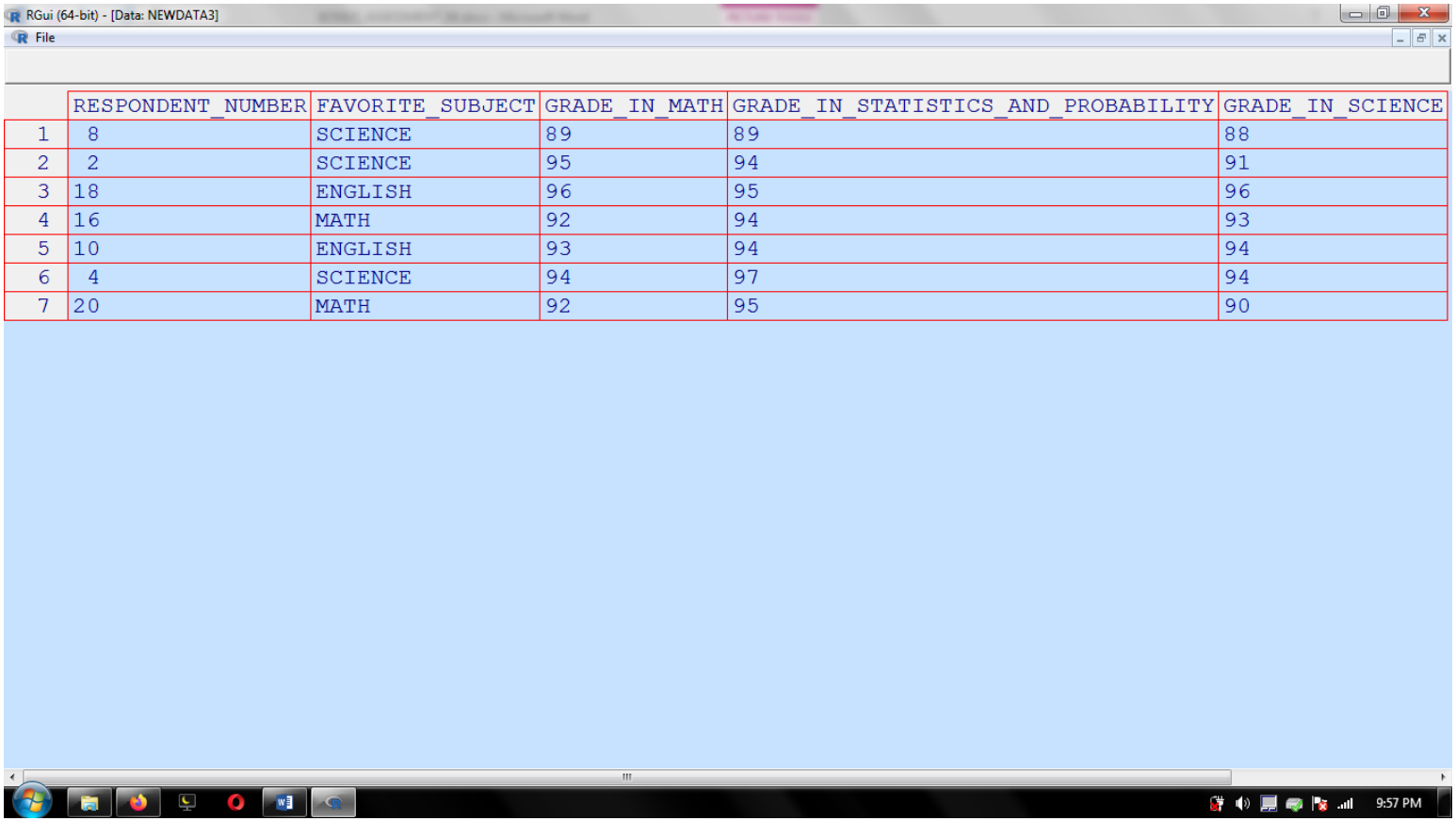
File

	RESPONDENT_NUMBER	GENDER	AGE	FAVORITE_SUBJECT	FAVORITE_COLOR	HEIGHT	WEIGHT	GRADE_IN_MATH
1	8	MALE	19	SCIENCE	YELLOW	167	45	89
2	2	FEMALE	19	SCIENCE	YELLOW	154	48	95
3	18	MALE	18	ENGLISH	BLUE	175	68	96
4	16	MALE	20	MATH	BLUE	174	49	92
5	10	FEMALE	19	ENGLISH	RED	162	54	93
6	4	MALE	19	SCIENCE	BLUE	165	45	94
7	20	MALE	18	MATH	BLUE	176	58	92



2. Include only those students whose favorite color are either red, blue or yellow.





	RESPONDENT_NUMBER	FAVORITE_SUBJECT	GRADE_IN_MATH	GRADE_IN_STATISTICS_AND_PROBABILITY	GRADE_IN_SCIENCE
1	8	SCIENCE	89	89	88
2	2	SCIENCE	95	94	91
3	18	ENGLISH	96	95	96
4	16	MATH	92	94	93
5	10	ENGLISH	93	94	94
6	4	SCIENCE	94	97	94
7	20	MATH	92	95	90

RGui (64-bit) - [Data: NEWDATA4]

File

	Respondent_Number	Favorite_Subject	Grade_in_Math	Grade_in_Statistics_and_Probability	Grade_in_Science
1	8	SCIENCE	89	89	88
2	2	SCIENCE	95	94	91
3	18	ENGLISH	96	95	96
4	16	MATH	92	94	93
5	10	ENGLISH	93	94	94
6	4	SCIENCE	94	97	94
7	20	MATH	92	95	90

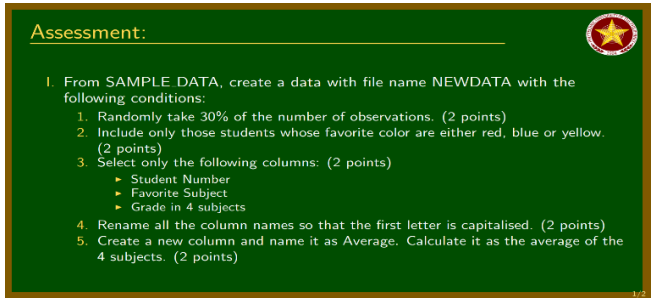
9:58 PM

3. Select only the following columns:

Student Number

Favorite Subject

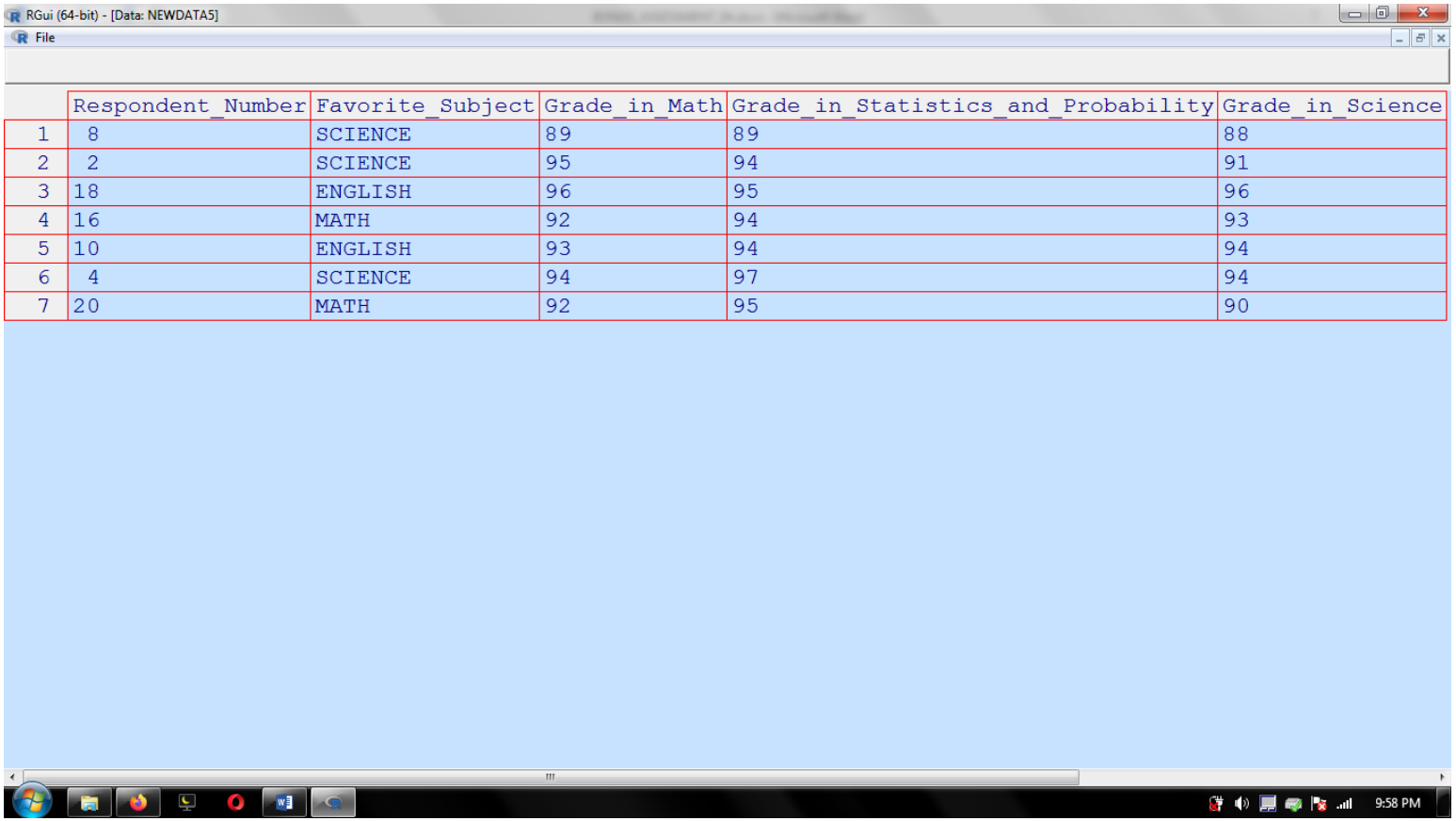
Grade in 4 subjects



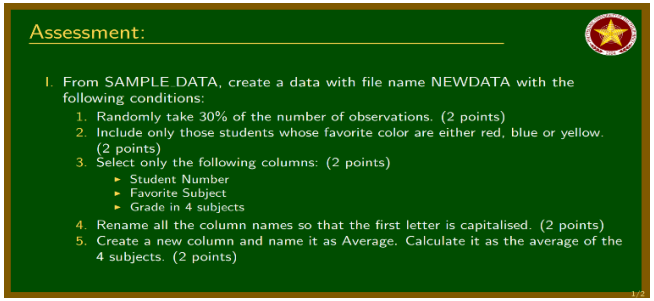
RGui (64-bit) - [Data: NEWDATA5]

File

	Respondent_Number	Favorite_Subject	Grade_in_Math	Grade_in_Statistics_and_Probability	Grade_in_Science
1	8	SCIENCE	89	89	88
2	2	SCIENCE	95	94	91
3	18	ENGLISH	96	95	96
4	16	MATH	92	94	93
5	10	ENGLISH	93	94	94
6	4	SCIENCE	94	97	94
7	20	MATH	92	95	90



4. Rename all the colum names so that the first letter is capitalized.



RGui (64-bit) - [Data: summary]

File

	Favorite_Subject	mean(Average, na.rm = T)	sd(Average, na.rm = T)
1	ENGLISH	94.50	1.414214
2	MATH	92.75	0.000000
3	SCIENCE	92.75	3.500000

9:58 PM

Create a new column and name it as Average.

Calculate it as the average of the 4 subjects.