

EXPLORATORY DATA ANALYSIS

LA1 Assignment

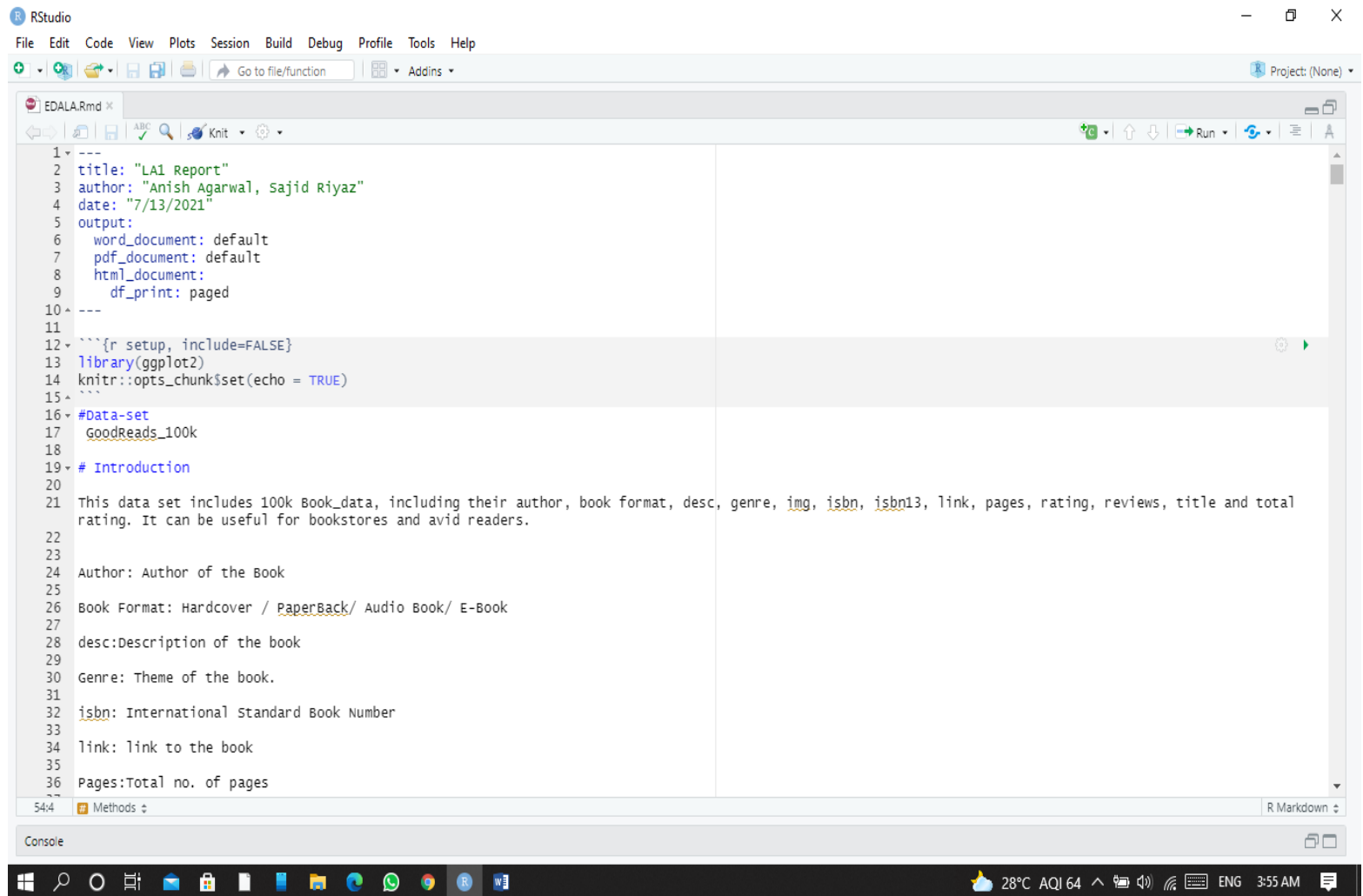
Anish Agarwal

1NT18IS028

Mohammad Sajid Riyaz

1NT18IS092

DATA SET: GOODREADS_100K_BOOK



The screenshot shows the RStudio IDE interface. The main editor window displays a markdown file named 'EDALA.Rmd'. The file content is as follows:

```
1 ---
2 title: "LA1 Report"
3 author: "Anish Agarwal, Sajid Riyaz"
4 date: "7/13/2021"
5 output:
6   word_document: default
7   pdf_document: default
8   html_document:
9     df_print: paged
10 ---
11
12 ```{r setup, include=FALSE}
13 library(ggplot2)
14 knitr::opts_chunk$set(echo = TRUE)
15 ```
16
17 #Data-set
18 GoodReads_100k
19
20 # Introduction
21
22 This data set includes 100k Book_data, including their author, book format, desc, genre, img, isbn, isbn13, link, pages, rating, reviews, title and total
23 rating. It can be useful for bookstores and avid readers.
24
25 Author: Author of the Book
26
27 Book Format: Hardcover / PaperBack/ Audio Book/ E-Book
28
29 desc:Description of the book
30
31 Genre: Theme of the book.
32
33 isbn: International Standard Book Number
34
35 link: link to the book
36
37 Pages:Total no. of pages
```

The RStudio interface includes a menu bar (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help), a toolbar with icons for file operations and running code, and a console window at the bottom. The system tray at the very bottom shows the date and time as 3:55 AM on 7/13/2021, along with system status icons.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALA.Rmd

```
37 # Pages: Total no. of pages
38 rating: Rating of the Book
39 Reviews: Reviews of the book
40 Title: Title of the book
41
42 Title: Title of the book
43
44 Total Rating : Overall Rating of the Book.
45
46 # Methods
47 1. Read the csv data file.
48 ```{r}
49 book_data = read.csv(file = file.choose(), header = TRUE)
50 ```
51 2. Display head of data
52 ```{r}
53 head(book_data)
54 ```
```

Description: df [6 x 13]

	i..author<chr>	bookformat<chr>
1	Laurence M. Hauptman	Hardcover
2	Charlotte Fiell, Emmanuelle Dirix	Paperback
3	Andy Anderson	Paperback
4	Carlotta R. Anderson	Hardcover
5	Jean Leveille	
6	Jeffrey Pfeffer	Hardcover

6 rows | 1-3 of 13 columns

```
55 3. Display first n rows specified.
56 ```{r}
57 head(book_data, n = 10)
58 ```
```

Console

28°C AQI 64 ENG 3:55 AM

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALA.Rmd

```
57 head(book_data, n = 10)
58 ```
```

Description: df [10 x 13]

	i..author<chr>	bookformat<chr>
1	Laurence M. Hauptman	Hardcover
2	Charlotte Fiell, Emmanuelle Dirix	Paperback
3	Andy Anderson	Paperback
4	Carlotta R. Anderson	Hardcover
5	Jean Leveille	
6	Jeffrey Pfeffer	Hardcover
7	Jeffrey Pfeffer	Paperback
8	Edward Joesting	Paperback
9	Nick Le Neve Walmsley	Paperback
10	B. Alan Wallace, Dalai Lama XIV	Hardcover

1-10 of 10 rows | 1-3 of 13 columns

```
59 4. Display tail of data.
60 ```{r}
61 tail(book_data)
62 ```
```

Description: df [6 x 13]

	i..author<chr>	bookformat<chr>
99995	Sarah Lean	Paperback
99996	Simon Monk	ebook
99997	Tracie L. Miller-Nobles, Brenda L. Mattison, Ella Mae Matsumura	Hardcover

Console

28°C Mostly cl... ENG 3:55 AM

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALARmd x

4. Display current data.

```
60 {r}
61 tail(book_data)
62
```

Description: df [6 x 13]

	i..author	bookformat
99995	Sarah Lean	Paperback
99996	Simon Monk	ebook
99997	Tracie L. Miller-Nobles,Brenda L. Mattison,Ella Mae Matsumura	Hardcover
99998	C. John Miller	Paperback
99999	Albert Marrin	Hardcover
100000	Marc E. Fitch	ebook

6 rows | 1-3 of 13 columns

```
63
64 5.Display n rows specified from bottom
65 {r}
66 tail(book_data,n=10)
67
```

Description: df [10 x 13]

	i..author	bookformat
99991	Ilja Gort	Hardcover
99992	Leonard Verduin	Paperback
99993	John Lescroart	Paperback
99994	Philip Hoare	Paperback
99995	Sarah Lean	Paperback
99996	Simon Monk	ebook

54:4 Methods R Markdown

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALARmd x

4. Display current data.

```
60 {r}
61 tail(book_data)
62
```

Description: df [6 x 13]

	i..author	bookformat
99995	Sarah Lean	Paperback
99996	Simon Monk	ebook
99997	Tracie L. Miller-Nobles,Brenda L. Mattison,Ella Mae Matsumura	Hardcover
99998	C. John Miller	Paperback
99999	Albert Marrin	Hardcover
100000	Marc E. Fitch	ebook

6 rows | 1-3 of 13 columns

```
63
64 5.Display n rows specified from bottom
65 {r}
66 tail(book_data,n=10)
67
```

Description: df [10 x 13]

	i..author	bookformat
99991	Ilja Gort	Hardcover
99992	Leonard Verduin	Paperback
99993	John Lescroart	Paperback
99994	Philip Hoare	Paperback
99995	Sarah Lean	Paperback
99996	Simon Monk	ebook

54:4 Methods R Markdown

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALARmd

72 7.Table command
73 {r}
74 table(book_data\$rating)
75 table(book_data\$pages)
76

0	1	1.31	1.33	1.5	1.67	1.68	1.71	1.75	1.83	1.85	1.92	1.93	1.95	2	2.05	2.07	2.1	2.17	2.18	2.2	2.23	2.25	2.26	2.27	2.28	2.29	2.3	2.31	2.33
1562	44	1	1	7	6	1	1	4	1	1	1	2	2	190	1	1	3	4	1	8	1	6	1	1	1	4	1	1	35
2.36	2.37	2.38	2.39	2.4	2.42	2.43	2.44	2.45	2.46	2.47	2.48	2.49	2.5	2.51	2.52	2.53	2.54	2.55	2.56	2.57	2.58	2.59	2.6	2.61	2.62	2.63	2.64	2.65	2.66
1	1	6	1	13	1	5	5	2	1	4	2	2	118	2	1	6	5	3	10	13	7	2	24	9	4	15	10	3	6
2.67	2.68	2.69	2.7	2.71	2.72	2.73	2.74	2.75	2.76	2.77	2.78	2.79	2.8	2.81	2.82	2.83	2.84	2.85	2.86	2.87	2.88	2.89	2.9	2.91	2.92	2.93	2.94	2.95	2.96
83	9	19	7	21	9	14	14	62	15	19	26	14	52	15	21	44	16	17	33	15	43	35	29	24	32	37	42	29	31
2.97	2.98	2.99	3	3.01	3.02	3.03	3.04	3.05	3.06	3.07	3.08	3.09	3.1	3.11	3.12	3.13	3.14	3.15	3.16	3.17	3.18	3.19	3.2	3.21	3.22	3.23	3.24	3.25	3.26
38	24	8	1097	14	35	52	42	49	54	59	74	72	80	125	65	131	138	107	94	181	144	124	234	151	162	142	129	294	156
3.27	3.28	3.29	3.3	3.31	3.32	3.33	3.34	3.35	3.36	3.37	3.38	3.39	3.4	3.41	3.42	3.43	3.44	3.45	3.46	3.47	3.48	3.49	3.5	3.51	3.52	3.53	3.54	3.55	3.56
188	158	267	213	201	201	568	216	255	307	239	371	255	430	290	326	395	402	403	365	382	403	349	1170	332	423	501	472	515	592
3.57	3.58	3.59	3.6	3.61	3.62	3.63	3.64	3.65	3.66	3.67	3.68	3.69	3.7	3.71	3.72	3.73	3.74	3.75	3.76	3.77	3.78	3.79	3.8	3.81	3.82	3.83	3.84	3.85	3.86
586	608	580	772	647	660	782	757	754	646	1248	794	801	894	998	785	959	855	1224	913	962	1015	991	1186	952	1035	1153	970	1063	1166
3.87	3.88	3.89	3.9	3.91	3.92	3.93	3.94	3.95	3.96	3.97	3.98	3.99	4	4.01	4.02	4.03	4.04	4.05	4.06	4.07	4.08	4.09	4.1	4.11	4.12	4.13	4.14	4.15	4.16
982	1179	1100	1054	1107	1103	1106	1134	1062	1041	995	977	797	3621	738	841	972	984	1039	1019	980	1025	996	956	1005	878	1044	964	857	808
4.17	4.18	4.19	4.2	4.21	4.22	4.23	4.24	4.25	4.26	4.27	4.28	4.29	4.3	4.31	4.32	4.33	4.34	4.35	4.36	4.37	4.38	4.39	4.4	4.41	4.42	4.43	4.44	4.45	4.46
1013	803	810	1039	781	782	700	692	955	612	636	557	701	634	510	486	946	405	416	409	361	453	319	511	319	308	343	339	300	226
4.47	4.48	4.49	4.5	4.51	4.52	4.53	4.54	4.55	4.56	4.57	4.58	4.59	4.6	4.61	4.62	4.63	4.64	4.65	4.66	4.67	4.68	4.69	4.7	4.71	4.72	4.73	4.74	4.75	4.76
243	186	161	922	122	153	136	111	138	145	182	102	116	186	77	64	93	82	46	32	286	45	43	45	66	21	33	22	160	15
4.77	4.78	4.79	4.8	4.81	4.82	4.83	4.84	4.85	4.86	4.87	4.88	4.89	4.9	4.91	4.92	4.93	4.95	5											
17	28	14	71	9	10	26	5	9	23	3	23	13	6	3	2	5	1	950											

54:4 Methods R Markdown

Console

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALARmd

1013 803 810 1039 781 782 700 692 955 612 636 557 701 634 510 486 946 405 416 409 361 453 319 511 319 308 343 339 300 226
4.47 4.48 4.49 4.5 4.51 4.52 4.53 4.54 4.55 4.56 4.57 4.58 4.59 4.6 4.61 4.62 4.63 4.64 4.65 4.66 4.67 4.68 4.69 4.7 4.71 4.72 4.73 4.74 4.75 4.76
243 186 161 922 122 153 136 111 138 145 182 102 116 186 77 64 93 82 46 32 286 45 43 45 66 21 33 22 160 15
4.77 4.78 4.79 4.8 4.81 4.82 4.83 4.84 4.85 4.86 4.87 4.88 4.89 4.9 4.91 4.92 4.93 4.95 5
17 28 14 71 9 10 26 5 9 23 3 23 13 6 3 2 5 1 950

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
7752	97	65	33	33	26	31	18	42	26	71	27	90	32	56	33	122	26	49	39	79	39	73	32	421	43	75	45	113	45
30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
102	75	2239	28	71	34	116	34	45	42	698	36	40	29	91	59	97	79	637	42	71	40	103	40	54	53	181	36	116	35
60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89
90	39	64	64	550	41	53	44	78	32	62	49	197	43	58	54	93	42	75	77	424	31	70	43	80	49	69	68	154	47
90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119
99	47	127	71	107	114	868	42	82	58	158	65	87	84	236	51	95	61	135	62	95	103	640	63	92	64	126	74	88	75
120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149
392	73	95	62	125	94	136	159	1430	61	96	54	161	76	81	82	295	68	88	62	149	73	114	139	1185	67	86	74	135	56
150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179
160	86	280	74	121	75	187	96	153	153	1411	73	113	112	146	96	99	113	372	63	135	87	174	101	141	160	1137	78	154	106
180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209
251	89	181	127	480	108	219	189	220	163	210	191	2453	79	119	84	211	91	146	108	721	84	146	100	186	105	168	122	1334	82
210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
182	96	182	85	155	112	475	106	207	129	270	120	220	153	2355	95	149	112	228	101	183	106	398	98	149	76	177	111	157	136
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269
1533	99	169	111	195	113	186	116	406	127	259	126	234	127	173	196	2196	94	153	89	210	102	171	118	486	86	143	91	127	81
270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299
164	112	1347	115	182	122	204	103	178	99	412	104	157	125	206	102	177	144	2166	95	173	104	190	113	170	120	401	103	145	94
300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329
259	89	164	119	1671	92	156	128	176	126	181	121	385	86	127	83	161	114	163	116	1890	84	158	104	201	108	178	90	312	74
330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359
130	72	120	78	125	110	1257	82	150	106	184	100	172	87	270	75	134	86	137	81	185	94	1339	66	112	68	144	84	127	79
360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389
297	65	98	62	98	69	105	82	937	71	117	80	158	85	133	103	202	70	102	64	103	70	89	80	1191	68	94	61	104	59
390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419
102	61	178	58	90	52	97	50	80	57	862	47	56	58	112	51	76	57	167	45	66	41	77	45	65	57	754	26	59	40
420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	4		

RStudio interface showing a script editor with R code and a console window. The script editor displays a large table of data (rows 570 to 1040) and a console window showing the output of the code, including a message: "reached getoption('max.print') -- omitted 357 entries".

```
570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599
18 10 31 10 13 18 190 10 17 7 28 9 12 8 35 11 19 12 12 9 13 11 120 13 17 7 27 9 13 11
600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629
66 3 16 6 19 10 15 9 117 8 15 9 19 10 21 9 21 3 7 4 18 6 14 9 113 10 10 5 17 5
630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659
9 9 27 4 8 7 16 9 16 11 112 5 11 5 8 6 11 11 25 3 16 7 6 3 10 11 66 10 5 9
660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689
11 7 12 7 24 3 6 4 8 6 5 8 110 5 6 4 11 4 9 6 15 11 10 3 13 8 6 7 59 4
690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719
4 3 13 6 5 12 11 6 5 13 4 10 4 83 5 10 2 7 10 7 7 13 5 2 7 6 3 10 3
720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749
65 5 9 4 5 1 3 6 9 7 3 1 10 4 7 3 70 1 6 3 8 4 3 2 15 5 7 1 5 1
750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 772 773 774 775 776 777 778 779 780 781
3 5 42 5 9 3 6 5 6 5 6 1 9 2 3 8 4 5 57 4 4 3 6 3 13 3 4 8 2
782 783 784 785 786 787 788 790 791 792 793 794 795 796 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813
4 4 44 4 5 5 3 6 3 10 2 3 1 6 4 2 51 4 2 2 4 3 4 2 8 2 4 4 6 1
814 815 816 817 818 819 820 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 842 843 844 845
1 3 27 3 8 6 8 4 1 8 3 4 6 5 3 7 3 32 2 2 4 10 1 3 3 5 2 4 5 4
846 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 876 877
4 19 2 8 2 3 2 5 1 7 4 6 1 8 3 2 4 46 2 1 1 3 4 5 1 5 3 1 2
878 879 880 881 883 884 885 886 887 888 889 890 891 892 893 894 896 898 900 901 902 904 906 907 908 909 910 912 913 914
2 3 21 2 3 5 3 6 1 11 2 7 2 3 1 4 23 1 4 1 1 5 4 1 3 3 14 2 2
915 916 917 918 919 920 921 922 923 924 925 927 928 929 930 931 933 935 936 937 938 939 940 941 942 943 944 945 946 947
1 1 2 2 1 2 1 2 1 2 1 15 1 1 1 1 1 2 8 3 1 3 2 1 3 2 19 2 1
948 949 950 952 953 954 956 957 958 959 960 962 963 964 965 966 967 968 969 970 973 974 975 976 977 978 979 980 981 982
5 1 2 3 1 2 3 4 4 3 27 2 2 2 1 3 1 1 1 2 1 2 1 18 3 1 1 2 4 2
983 984 986 987 988 989 990 991 992 993 994 998 1000 1002 1004 1005 1006 1007 1008 1010 1012 1014 1015 1017 1018 1019 1020 1022 1024 1025
1 1 1 2 6 2 2 3 10 2 2 2 2 1 1 5
1028 1029 1030 1031 1032 1035 1036 1037 1039 1040
1 1 1 1 5 2 2 1 1 5
[ reached getoption("max.print") -- omitted 357 entries ]

77
78 8.Determine the structure of data
79- '{r}'
80 str(book_data)
81- ''
```

RStudio interface showing a script editor with R code and a console window. The script editor displays a large table of data (rows 1028 to 1040) and a console window showing the output of the code, including a message: "reached getoption('max.print') -- omitted 357 entries".

```
1028 1029 1030 1031 1032 1035 1036 1037 1039 1040
1 1 1 1 5 2 2 1 1 5
[ reached getoption("max.print") -- omitted 357 entries ]

77
78 8.Determine the structure of data
79- '{r}'
80 str(book_data)
81- ''

'data.frame': 100000 obs. of 13 variables:
 $ i..author : chr "Laurence M. Hauptman" "Charlotte Fiell,Emmanuelle Dirix" "Andy Anderson" "Carlotta R. Anderson" ...
 $ bookformat : chr "Hardcover" "Paperback" "Paperback" "Hardcover" ...
 $ desc : chr "Reveals that several hundred thousand Indians were affected by the Civil war and that twenty thousand Indians e"| __truncated__
 "Fashion Sourcebook - 1920s is the first book in a brand-new series by Fiell Publishing that documents comprehen"| __truncated__ "The seminal history and
 analysis of the Hungarian Revolution and the workers' councils, perhaps the single most"| __truncated__ "\"All-American Anarchist\" chronicles the life
 and work of Joseph A. Labadie (1850-1933), Detroit's prominent l"| __truncated__ ...
 $ genre : chr "History,Military History,Civil war,American History,American Civil war,Nonfiction,North American Hi...,American"| __truncated__
 "Couture,Fashion,Historical,Art,Nonfiction" "Politics,History" "Labor,History" ...
 $ img : chr "https://i.gr-assets.com/images/S/compressed.photo.goodreads.com/books/13877387651/1001053.jpg" "https://i.gr-
 assets.com/images/S/compressed.photo.goodreads.com/books/14210114971/10010552.jpg" "https://i.gr-
 assets.com/images/S/compressed.photo.goodreads.com/books/13481177081/1001077.jpg" "https://i.gr-
 assets.com/images/S/compressed.photo.goodreads.com/books/13564612141/1001079.jpg" ...
 $ isbn : chr "002914180x" "1906863482" "948984147" "814327079" ...
 $ isbn13 : chr "9.78E+12" "9.78E+12" "9.78E+12" "9.78E+12" ...
 $ link : chr "https://goodreads.com/book/show/1001053.Between_Two_Fires" "https://goodreads.com/book/show/10010552-fashion-sourcebook-1920s"
 "https://goodreads.com/book/show/1001077.Hungary_56" "https://goodreads.com/book/show/1001079.All_American_Anarchist" ...
 $ pages : int 0 576 124 324 177 368 304 353 160 256 ...
 $ rating : num 3.52 4.51 4.15 3.83 4 3.73 3.65 3.93 5 4.1 ...
 $ reviews : int 5 6 2 1 1 7 1 2 2 7 ...
 $ title : chr "Between Two Fires: American Indians in the civil war" "Fashion Sourcebook 1920s" "Hungary 56" "All-American Anarchist: Joseph A.
 Labadie and the Labor Movement" ...
 $ totalratings: int 33 41 26 6 1 119 20 15 3 133 ...

82 9.Summarising the data
83- '{r}'
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

EDALARmd x

Labadie and the Labor Movement" ...
\$ totalratings: int 33 41 26 6 1 119 20 15 3 133 ...

```
82 9. Summarizing the data
83 ~~~{r}
84 summary(book_data)
85 ~~~
```

i..author	bookformat	desc	genre	img	isbn	isbn13
Length:100000	Length:100000	Length:100000	Length:100000	Length:100000	Length:100000	Length:100000
Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character

link	pages	rating	reviews	title	totalratings
Length:100000	Min. : 0	Min. :0.000	Min. : 0.0	Length:100000	Min. : 0
Class :character	1st Qu.: 135	1st Qu.:3.660	1st Qu.: 3.0	Class :character	1st Qu.: 31
Mode :character	Median : 240	Median :3.910	Median : 15.0	Mode :character	Median : 146
	Mean : 255	Mean :3.833	Mean : 181.5		Mean : 2991
	3rd Qu.: 336	3rd Qu.:4.140	3rd Qu.: 67.0		3rd Qu.: 744
	Max. :70000	Max. :5.000	Max. :158776.0		Max. :3819326

```
86 10. Displaying Dimension of the data
87 ~~~{r}
88 dim(book_data)
89 ~~~
```

```
[1] 100000 13
```

```
90 11. Displaying length of attack column
91 ~~~{r}
92 length(book_data$rating)
93 ~~~
```

54:4 Methods R Markdown

Console

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

EDALARmd x

Mean : 255	Mean :3.833	Mean : 181.5	Mean : 2991
3rd Qu.: 336	3rd Qu.:4.140	3rd Qu.: 67.0	3rd Qu.: 744
Max. :70000	Max. :5.000	Max. :158776.0	Max. :3819326

```
86 10. Displaying Dimension of the data
87 ~~~{r}
88 dim(book_data)
89 ~~~
```

```
[1] 100000 13
```

```
90 11. Displaying length of attack column
91 ~~~{r}
92 length(book_data$rating)
93 ~~~
```

```
[1] 100000
```

```
94 12. Displaying column names of data
95 ~~~{r}
96 colnames(book_data)
97 ~~~
```

[1] "i..author"	"bookformat"	"desc"	"genre"	"img"	"isbn"	"isbn13"	"link"	"pages"
[10] "rating"	"reviews"	"title"	"totalratings"					

```
98 14. Displaying type of some data structure in the data
99 ~~~{r}
100 typeof(book_data$pages)
101 ~~~
102 ~~~
```

```
[1] "integer"
```

54:4 Methods R Markdown

Console

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALARmd x

Knit

```
[1] "i..author" "bookformat" "desc" "genre" "img" "isbn" "isbn13" "link" "pages"
[10] "rating" "reviews" "title" "totalratings"
```

98
99 14.Displaying type of some data structure in the data
100 ```{r}
101 typeof(book_data\$pages)
102 ```

```
[1] "integer"
```

103 15.List of variables present in book data
104 ```{r}
105 ls(book_data)
106 ```

```
[1] "bookformat" "desc" "genre" "i..author" "img" "isbn" "isbn13" "link" "pages"
[10] "rating" "reviews" "title" "totalratings"
```

107 16.Display 1st row and all columns of data frame
108 ```{r}
109 book_data[1,]
110 ```

Description: df [1 x 13]

i..author	bookformat
1 Laurence M. Hauptman	Hardcover

1 row | 1-3 of 13 columns

54:4 Methods R Markdown

28°C Mostly cl... ENG 3:57 AM

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALARmd x

Knit

```
108 ```{r}
109 book_data[1,]
110 ```
```

Description: df [1 x 13]

i..author	bookformat
1 Laurence M. Hauptman	Hardcover

1 row | 1-3 of 13 columns

111 17.Display all rows and 1st column of data frame
112 ```{r}
113 book_data[,1]
114 ```

```
[1] "Laurence M. Hauptman"
[2] "Charlotte Fiell,Emmanuelle Dirix"
[3] "Andy Anderson"
[4] "Carlotta R. Anderson"
[5] "Jean Leveille"
[6] "Jeffrey Pfeffer"
[7] "Jeffrey Pfeffer"
[8] "Edward Joesting"
[9] "Nick Le Neve Walmsley"
[10] "B. Alan Wallace,Dalai Lama XIV"
[11] "Mark Verman"
[12] "Brian Morris"
[13] "Graham Purchase"
[14] "Richard Layard"
[15] "Graham Purchase"
[16] "Philip Mansel"
[17] "Umberto Eco"
[18] "Helen Fremont"
```

54:4 Methods R Markdown

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALA.Rmd

```
[46] "Janice G. Raymond"
[47] "Brian G. Hedges,Donald S. Whitney"
[48] "Richard L. Kagan"
[49] "Richard L. Kagan"
[50] "Louis L'Amour"
[51] "Jeanne Betancourt,Robert S. Brown"
[52] "Jeanne Betancourt,Robert S. Brown"
[53] "Jeanne Betancourt"
[54] "Jeanne Betancourt"
[55] "Brenda Shoshanna"
[56] "Jeanne Betancourt"
[57] "Jeanne Betancourt"
[58] "Jeanne Betancourt,Paul Bachem"
[59] "Chuck Dixon,Scott McDaniel,Karl Story"
[60] "Jeanne Betancourt,Vivien Kubbos"
[61] "Jeanne Betancourt"
[62] "Jeanne Betancourt,Paul Bachem"
[63] "Jeanne Betancourt"
[64] "Carrie Bickner"
[65] "Michael Camille"
[66] "A.R. Jarvis"
[67] "Richard Wiseman"
[68] "Eliezer Yudkowsky"
[69] "Megan Derr,Sasha L. Miller"
[70] "Sunita Apte"
[71] "Jaiya John"
[72] "John C. Maxwell"
[73] "Stephen L. Nelson"
[74] "Rudolf Steiner"
[75] "Christian Zellweger"
[76] "Rob Boyle,Brian Cross,J.M. Hardy,Olivia Hill,Ken Horner,Evan Jamieson,James Kneivitt,T.S. Luikart,Phil Masters,Steven Mohan,Malcolm Sheppard,John Snead,Tobias Wolter,Michelle Lyons"
[77] "Sandstorm Productions"
[78] "Nicole Monet"
[79] "C.J. Dennis"
[80] "Robert Drews"
[81] "Rodney Castleden"
```

54:4 Methods

R Markdown

Console

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALA.Rmd

```
[997] "Alfred wainwright"
[998] "Alfred wainwright,Derry Brabbs"
[999] "Alfred wainwright"
[1000] "Beth Bernobich"
[ reached getoption("max.print") -- omitted 99000 entries ]

115 18.Display data in 2nd row and 3rd column of the data frame
116 ```{r}
117 book_data[2,3]
118 ```

[1] "Fashion Sourcebook - 1920s is the first book in a brand-new series by Fiell Publishing that documents comprehensively the seasonal fashion styles of the 20th century, decade by decade. Sumptuously illustrated with over 600 original photographs, drawings and prints, this title is a must-have reference work for not only students of fashion, but for all fashionistas. Fashion Sourcebook - 1920s focuses on the Art Deco period with its beautiful beaded dresses, cloche hats and t-bar shoes as worn by the fashionable flappers and the \"bright young things\" of the time. An accompanying introduction outlines the major themes within fashion during this period and introduces its most famous designers and assesses their creative contributions. Text in English, French & German. Also Available: Fashion Sourcebook - 1930s ISBN: 9781906863586 24.95\"

119 19.Display 1st and 2nd row and all columns
120 ```{r}
121 book_data[1:2,]
122 ```

Description: df [2 x 13]



|   | i.author                         | bookformat |
|---|----------------------------------|------------|
| 1 | Laurence M. Hauptman             | Hardcover  |
| 2 | Charlotte Fiell,Emmanuelle Dirix | Paperback  |



2 rows | 1-3 of 13 columns

123 20.Display all rows and first 3 columns
124 ```{r,eval=F}
125 ```
```

54:4 Methods

R Markdown

Console

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

EDALA.Rmd x

2 rows | 1-3 of 13 columns

```
123
124 20.Display all rows and first 3 columns
125 ```{r,eval=F}
126 book_data[,3]
127 ```
128
129 22.Display sum of pages column
130 ```{r,eval=F}
131 sum(book_data[9])
132 ```
133
134 [1] 25501024
135
136 23.Display the maximum value of the rating column
137 ```{r,eval=F}
138 max(book_data[9])
139 ```
140
141 [1] 70000
142
143 24.Display the minimum value of the rating column
144 ```{r,eval=F}
145 min(book_data[9])
146 ```
147
148 [1] 0
149
150 25.Attaching book data
151 ```{r}
152 attach(book_data)
153 ```
```

54:4 Methods R Markdown

Console

28°C Mostly cl... ENG 3:57 AM

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

EDALA.Rmd x

```
144 25.Attaching book data
145 ```{r}
146 attach(book_data)
147 ```
148
149 26.Now we can use variables inside book data
150 ```{r,eval=F}
151 min(pages)
152 tail(rating)
153 ```
154
155 [1] 0
156 [1] 3.97 4.07 4.05 4.27 3.63 3.83
157
158 27.sorting pages variable in ascending order
159 ```{r,eval=F}
160 sort(pages)
161 ```
162
163 28.Sorting pages variable in descending order
164 ```{r,eval=F}
165 sort(pages,decreasing = TRUE)
166 ```
167
168 29.Detaching book data
169 ```{r}
170 detach(book_data)
171 ```
172
173 30.Using with operator to use variables inside data
174 ```{r,eval=F}
175 with(book_data,rating)
176 ```
177
178 31.Finding median of data
179 ```{r}
180 median(rating)
181 ```
```

54:4 Methods R Markdown

Console

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

EDALA.Rmd

```
172 31. Finding median of data
173 ```{r}
174 median(book_data$rating)
175 ```
176 [1] 3.91
177
178
179 32. Finding mean of data
180 ```{r}
181 mean(book_data$rating)
182 ```
183 [1] 3.833055
184
185 33. Finding standard deviation of data
186 ```{r}
187 sd(book_data$pages)
188 ```
189 [1] 367.9136
190
191 34. Finding variance of data
192 ```{r,eval=F}
193 var(book_data$pages)
194 ```
195 [1] 135360.4
196
197 35. Order the price column in ascending order
198 ```{r,eval=F}
199 order(book_data$price)
200 ```
```

Console

R Markdown

28°C Mostly cl... 3:57 AM

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

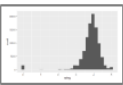
Project: (None)

EDALA.Rmd

```
194 35. Order the price column in ascending order
195 ```{r,eval=F}
196 order(book_data$price)
197 ```
198 [1] 135360.4
199
200 36. Order the rating column in descending order
201 ```{r,eval=F}
202 order(book_data$rating, decreasing = TRUE)
203 ```
204
205 37. Rank of rating column
206 ```{r,eval=F}
207 rank(book_data$pages)
208 ```
209
210 38. Rank of rating column with average as tie breaker
211 ```{r,eval=F}
212 rank(book_data$rating, ties.method = "average")
213 ```
214
215 39. Histogram
216 ```{r,eval=F}
217 ggplot(book_data, aes(x = rating)) + geom_histogram()
218 ```
```

Console

R Markdown



20000

