Question 8

a) I read in the data using the read.csv() function:

   a.
```
# a)
college = read.csv("../College.csv")
```

b) After using the book's code to get the row names, here is the fixed data:

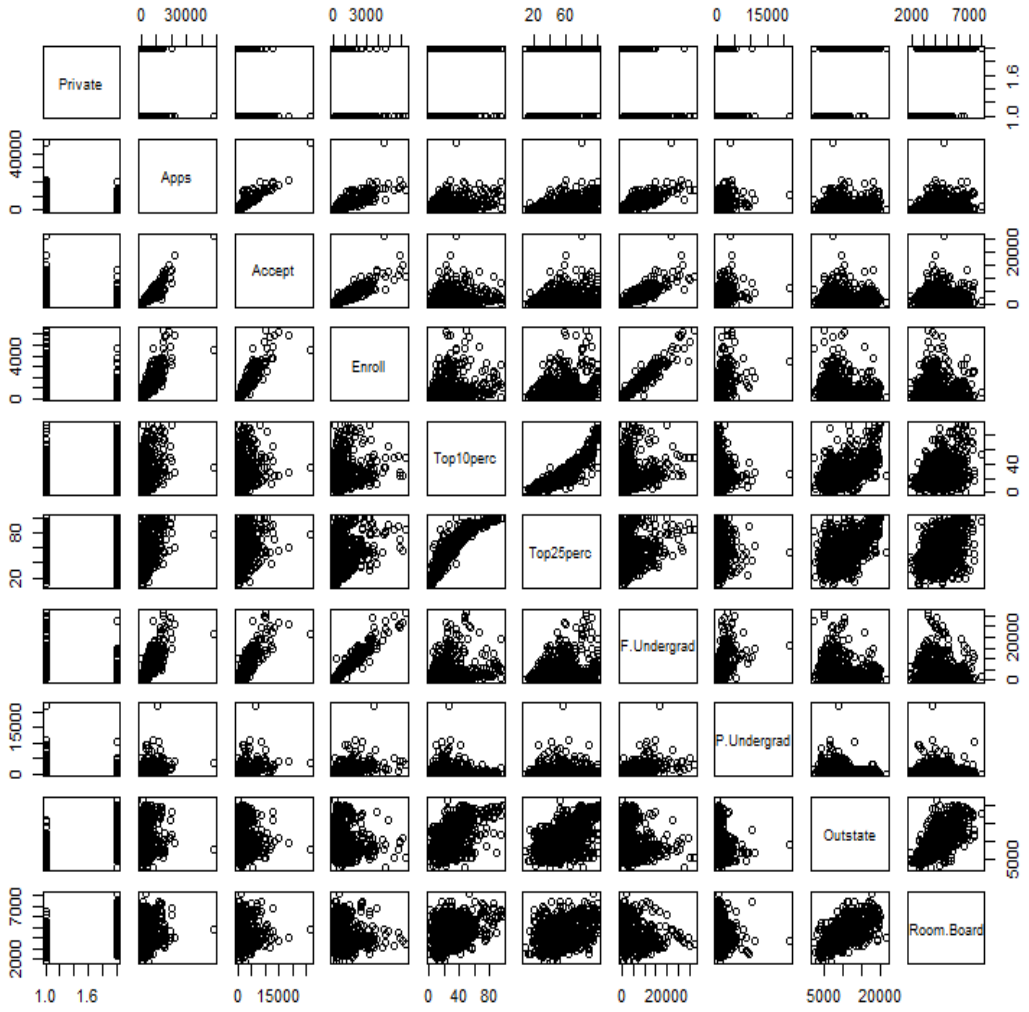
   a.

	row.names	Private	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
1	Abilene Christian University	Yes	1660	1232	721	23	52	2885	537	7440	3300	450	2200	70	78	18.1	12	7041	60
2	Adelphi University	Yes	2186	1924	512	16	29	2683	1227	12280	6450	750	1500	29	30	12.2	16	10527	56
3	Adrian College	Yes	1428	1097	336	22	50	1036	99	11250	3750	400	1165	53	66	12.9	30	8735	54
4	Agnes Scott College	Yes	417	349	137	60	89	510	63	12960	5450	450	875	92	97	7.7	37	19016	59
5	Alaska Pacific University	Yes	193	146	55	16	44	249	869	7560	4120	800	1500	76	72	11.9	2	10922	15
6	Albertson College	Yes	587	479	158	38	62	678	41	13500	3335	500	675	67	73	9.4	11	9727	55
7	Albertus Magnus College	Yes	353	340	103	17	45	416	230	13290	5720	500	1500	90	93	11.5	26	8861	63
8	Albion College	Yes	1899	1720	489	37	68	1594	32	13868	4826	450	850	89	100	13.7	37	11487	73
9	Albright College	Yes	1038	839	227	30	63	973	306	15595	4400	300	500	79	84	11.3	23	11644	80
10	Alderson-Broaddus College	Yes	582	498	172	21	44	799	78	10468	3380	660	1800	40	41	11.5	15	8991	52
11	Alfred University	Yes	1732	1425	472	37	75	1830	110	16548	5406	500	600	82	88	11.3	31	10932	73
12	Allegheny College	Yes	2652	1900	484	44	77	1707	44	17080	4440	400	600	73	91	9.9	41	11711	76
13	Allentown Coll. of St. Francis de Sales	Yes	1179	780	290	38	64	1130	638	9690	4785	600	1000	60	84	13.3	21	7940	74
14	Alma College	Yes	1267	1080	385	44	73	1306	28	12572	4552	400	400	79	87	15.3	32	9305	68
15	Alverno College	Yes	494	313	157	23	46	1317	1235	8352	3640	650	2449	36	69	11.1	26	8127	55
16	American International College	Yes	1420	1093	220	9	22	1018	287	8700	4780	450	1400	78	84	14.7	19	7355	69
17	Amherst College	Yes	4302	992	418	83	96	1593	5	19760	5300	660	1598	93	98	8.4	63	21424	100
18	Anderson University	Yes	1216	908	423	19	40	1819	281	10100	3520	550	1100	48	61	12.1	14	7994	59
19	Andrews University	Yes	1130	704	322	14	23	1586	326	9996	3090	900	1320	62	66	11.5	18	10908	46
20	Angelo State University	No	3540	2001	1016	24	54	4190	1512	5130	3592	500	2000	60	62	23.1	5	4010	34
21	Antioch University	Yes	713	661	252	25	44	712	23	15476	3336	400	1100	69	82	11.3	35	42926	48
22	Appalachian State University	No	7313	4664	1910	20	63	9940	1035	6806	2540	96	2000	83	96	18.3	14	5854	70
23	Aquinas College	Yes	619	516	219	20	51	1251	767	11208	4124	350	1615	55	65	12.7	25	6584	65

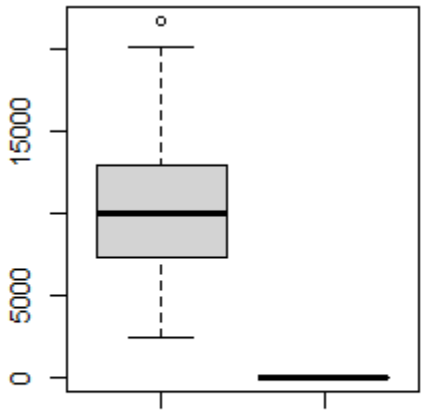
c) Here is the numerical summary of the data using the summary() function:

   i.
```
   Private                Apps           Accept          Enroll       Top10perc
 Length:777         Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
 Class :character   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
 Mode  :character   Median : 1558   Median : 1110   Median : 434   Median :23.00
                    Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.56
                    3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00
                    Max.   :48094   Max.   :26330   Max.   :6392   Max.   :96.00
   Top25perc      F.Undergrad     P.Undergrad       Outstate     Room.Board
 Min.   :  9.0   Min.   :  139   Min.   :    1.0   Min.   : 2340   Min.   :1780
 1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320   1st Qu.:3597
 Median : 54.0   Median : 1707   Median :  353.0   Median : 9990   Median :4200
 Mean   : 55.8   Mean   : 3700   Mean   :  855.3   Mean   :10441   Mean   :4358
 3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925   3rd Qu.:5050
 Max.   :100.0   Max.   :31643   Max.   :21836.0   Max.   :21700   Max.   :8124
     Books           Personal         PhD            Terminal       S.F.Ratio
 Min.   :  96.0   Min.   : 250   Min.   :  8.00   Min.   : 24.0   Min.   : 2.50
 1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00   1st Qu.: 71.0   1st Qu.:11.50
 Median : 500.0   Median :1200   Median : 75.00   Median : 82.0   Median :13.60
 Mean   : 549.4   Mean   :1341   Mean   : 72.66   Mean   : 79.7   Mean   :14.09
 3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00   3rd Qu.: 92.0   3rd Qu.:16.50
 Max.   :2340.0   Max.   :6800   Max.   :103.00   Max.   :100.0   Max.   :39.80
  perc.alumni        Expend        Grad.Rate
 Min.   : 0.00   Min.   : 3186   Min.   : 10.00
 1st Qu.:13.00   1st Qu.: 6751   1st Qu.: 53.00
 Median :21.00   Median : 8377   Median : 65.00
 Mean   :22.74   Mean   : 9660   Mean   : 65.46
 3rd Qu.:31.00   3rd Qu.:10830   3rd Qu.: 78.00
 Max.   :64.00   Max.   :56233   Max.   :118.00
>
```

ii. Result from pairs() function:



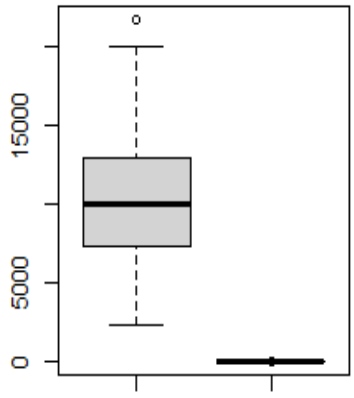
iii. Result from Boxplot of Outstate / Private:



Outstate | Private

iv.

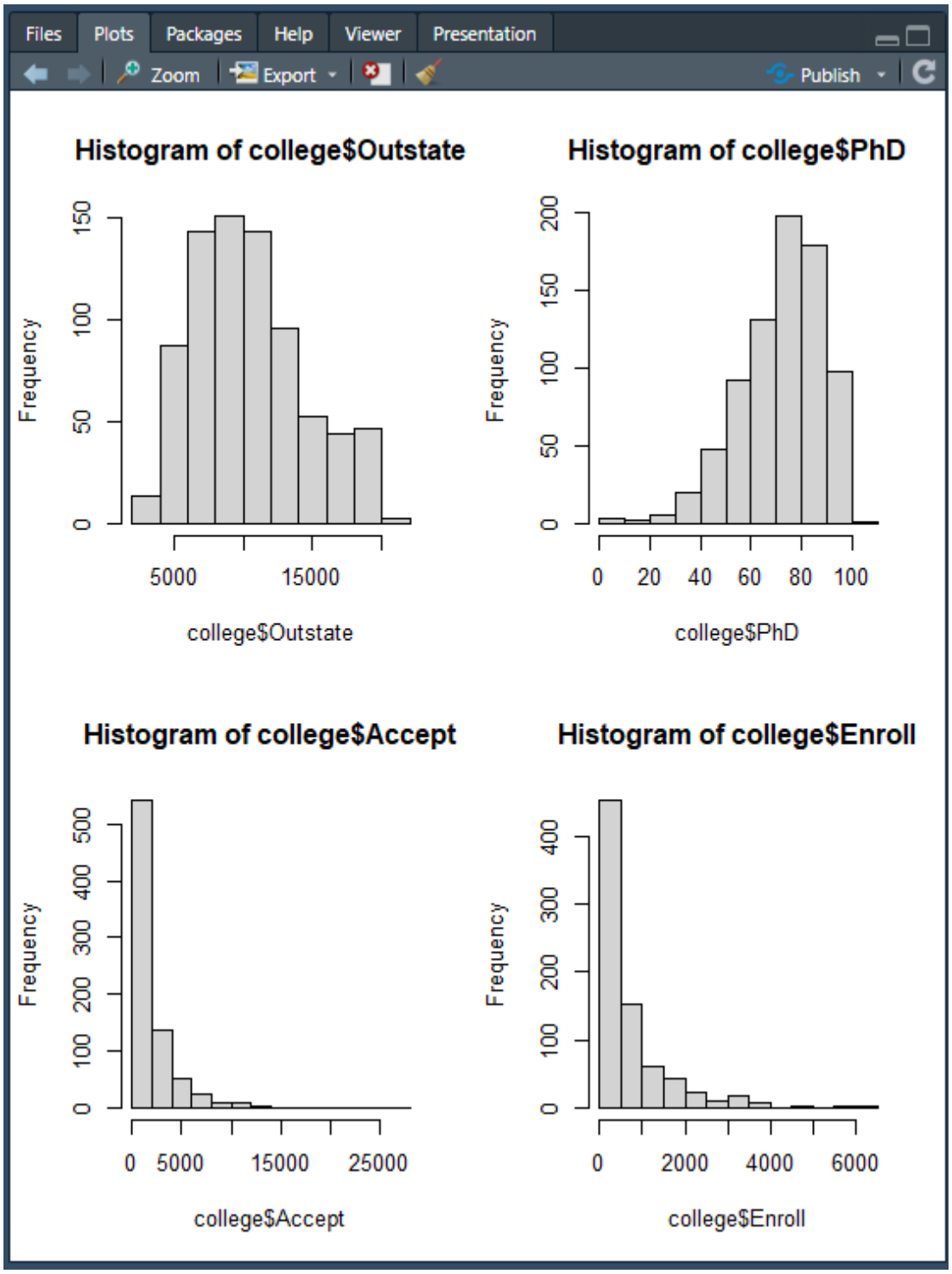
```
> pairs(college[,1:10])
> #   iii.
> boxplot(college$Outstate, college$Private,
xlab="Outstate | Private")
> #   iv.
> Elite=rep("No", nrow(college))
> Elite[college$Top10perc>50]="Yes"
> Elite=as.factor(Elite)
> college=data.frame(college,Elite)
> print("--------------------")
[1] "--------------------"
> print(summary(Elite))
 No Yes
699  78
> boxplot(college$Outstate, college$Elite, x
lab="Outstate | Elite")
>
```



Outstate | Elite

1. The No / Yes printed in the console dialog is the number of Colleges identified as Elite

v. Here are a few of the histograms I created from the data:



Question 10

a) There are 506 rows and 14 columns, the rows are different suburbs in Boston and the columns all represent attributes about those suburbs

d) Here are the Summaries of the requested columns:

```
[1] "------"
> print(summary(Boston$crim))
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
 0.00632  0.08204  0.25651  3.61352  3.67708 88.97620
> print("------")
[1] "------"
> print(summary(Boston$tax))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  187.0   279.0   330.0   408.2   666.0   711.0
> print("------")
[1] "------"
> print(summary(Boston$ptratio))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.60   17.40   19.05   18.46   20.20   22.00
>
```

i. The crime rates have a huge range, spanning .006 to 88.98. The highest suburbs are 381, 419, and 406.

ii. Tax rates have a fairly large range as well, from 187 to 711. The distributions for both tax and P.T.Ratio are fairly evenly spread.

iii. P.T.Ratio does not have a very large range, only 12.6 to 22

e) According to the data, there are 35 suburbs by the river:

```
[1] "Suburbs by river: "
> print(sum(Boston$chas))
[1] 35
```

f) The median pupil-teacher ratio is 19.05:

```
[1] "Median ptRatio:"
> print(median(Boston$ptratio))
[1] 19.05
```

g) Suburb #42 has the lowest median value of owner-occupied homes. It has a fairly low lstat at 4.84 compared to the average of 12.65:

```
> print(min(Boston$age))
[1] 2.9
> print(mean(Boston$lstat))
[1] 12.65306
>
```

h)

i. There are 64 suburbs that have more than 7 rooms per dwelling:

```
> print(dim(bdf[bdf$rm>7,]))
[1] 64 14
```

ii. There are 13 suburbs that have more than 8 rooms per dwelling:

```
> print(dim(bdf[bdf$rm>8,]))
[1] 13 14
```

iii.

```
> print(bdf[bdf$rm>8,])
       crim zn indus chas    nox    rm  age    dis rad tax ptratio  black lstat
98  0.12083  0  2.89    0 0.4450 8.069 76.0 3.4952   2 276    18.0 396.90  4.21
164 1.51902  0 19.58    1 0.6050 8.375 93.9 2.1620   5 403    14.7 388.45  3.32
205 0.02009 95  2.68    0 0.4161 8.034 31.9 5.1180   4 224    14.7 390.55  2.88
225 0.31533  0  6.20    0 0.5040 8.266 78.3 2.8944   8 307    17.4 385.05  4.14
226 0.52693  0  6.20    0 0.5040 8.725 83.0 2.8944   8 307    17.4 382.00  4.63
227 0.38214  0  6.20    0 0.5040 8.040 86.5 3.2157   8 307    17.4 387.38  3.13
233 0.57529  0  6.20    0 0.5070 8.337 73.3 3.8384   8 307    17.4 385.91  2.47
234 0.33147  0  6.20    0 0.5070 8.247 70.4 3.6519   8 307    17.4 378.95  3.95
254 0.36894 22  5.86    0 0.4310 8.259  8.4 8.9067   7 330    19.1 396.90  3.54
258 0.61154 20  3.97    0 0.6470 8.704 86.9 1.8010   5 264    13.0 389.70  5.12
263 0.52014 20  3.97    0 0.6470 8.398 91.5 2.2885   5 264    13.0 386.86  5.91
268 0.57834 20  3.97    0 0.5750 8.297 67.0 2.4216   5 264    13.0 384.54  7.44
365 3.47428  0 18.10    1 0.7180 8.780 82.9 1.9047  24 666    20.2 354.55  5.29
```

1. They all roughly have an average ptratio, and the ages are all very high, except for suburb 254