Iterative Generation of Frequency Distributions

Robert Schnitman

November 21, 2019

The purpose of this document is to demonstrate the production of a multi-paged report of frequency distributions via R Markdown. What follows is a cross-tabulation between the color and quality of diamonds by their depth level. The data used is `diamonds` from the R library `ggplot2`.

At a high level, we perform the following to generate the tables in this report:

1. Load table-styling and data management libraries (`knitr`, `kableExtra`, `tidyverse`, and `magrittr`). This step assumes that `tinytex` has been installed in RStudio and that `tinytex::install_tinytex()` has been executed.
2. Create a function that combines the counts and row-wise percentages in their own dataset (more specifically, a `data frame`).
3. Write a function that generalizes the preferred table style: counts and row-wise percentages for each row-group, along with computed means of the quality level.
4. Apply steps 2 and 3 above for each depth level via a `for` loop (i.e. "for each depth level, generate the cross-tabulation between the diamonds' color and quality category").

The source code is located at https://github.com/robertschnitman/RS_Reports/blob/master/FD/FD.rmd.

Table 1: Clarity: SI2

Color	1 = Fair	2 = Good	3 = Very Good	4 = Premium	5 = Ideal	Total	Mean Quality Level
D	56 4%	223 16%	314 23%	421 31%	356 26%	1,370 100%	3.58
E	78 5%	202 12%	445 26%	519 30%	469 27%	1,713 100%	3.64
F	89 6%	201 12%	343 21%	523 33%	453 28%	1,609 100%	3.65
G	80 5%	163 11%	327 21%	492 32%	486 31%	1,548 100%	3.74
H	91 6%	158 10%	343 22%	521 33%	450 29%	1,563 100%	3.69
I	45 5%	81 9%	200 22%	312 34%	274 30%	912 100%	3.76
J	27 6%	53 11%	128 27%	161 34%	110 23%	479 100%	3.57

Table 2: Clarity: SI1

Color	1 = Fair	2 = Good	3 = Very Good	4 = Premium	5 = Ideal	Total	Mean Quality Level
D	58 3%	237 11%	494 24%	556 27%	738 35%	2,083 100%	3.81
E	65 3%	355 15%	626 26%	614 25%	766 32%	2,426 100%	3.68
F	83 4%	273 13%	559 26%	608 29%	608 29%	2,131 100%	3.65
G	69 3%	207 10%	474 24%	566 29%	660 33%	1,976 100%	3.78
H	75 3%	235 10%	547 24%	655 29%	763 34%	2,275 100%	3.79
I	30 2%	165 12%	358 25%	367 26%	504 35%	1,424 100%	3.81
J	28 4%	88 12%	182 24%	209 28%	243 32%	750 100%	3.73

Table 3: Clarity: VS1

Color	1 = Fair	2 = Good	3 = Very Good	4 = Premium	5 = Ideal	Total	Mean Quality Level
D	5 1%	43 6%	175 25%	131 19%	351 50%	705 100%	4.11
E	14 1%	89 7%	293 23%	292 23%	593 46%	1,281 100%	4.06
F	33 2%	132 10%	293 21%	290 21%	616 45%	1,364 100%	3.97
G	45 2%	152 7%	432 20%	566 26%	953 44%	2,148 100%	4.04
H	32 3%	77 7%	257 22%	336 29%	467 40%	1,169 100%	3.97
I	25 3%	103 11%	205 21%	221 23%	408 42%	962 100%	3.92
J	16 3%	52 10%	120 22%	153 28%	201 37%	542 100%	3.87

Table 4: Clarity: VS2

Color	1 = Fair	2 = Good	3 = Very Good	4 = Premium	5 = Ideal	Total	Mean Quality Level
D	25 1%	104 6%	309 18%	339 20%	920 54%	1,697 100%	4.19
E	42 2%	160 6%	503 20%	629 25%	1,136 46%	2,470 100%	4.08
F	53 2%	184 8%	466 21%	619 28%	879 40%	2,201 100%	3.95
G	45 2%	192 8%	479 20%	721 31%	910 39%	2,347 100%	3.96
H	41 2%	138 8%	376 23%	532 32%	556 34%	1,643 100%	3.87
I	32 3%	110 9%	274 23%	315 27%	438 37%	1,169 100%	3.87
J	23 3%	90 12%	184 25%	202 28%	232 32%	731 100%	3.73

Table 5: Clarity: VVS2

Color	1 = Fair	2 = Good	3 = Very Good	4 = Premium	5 = Ideal	Total	Mean Quality Level
D	9 2%	25 5%	141 25%	94 17%	284 51%	553 100%	4.12
E	13 1%	52 5%	298 30%	121 12%	507 51%	991 100%	4.07
F	10 1%	50 5%	249 26%	146 15%	520 53%	975 100%	4.14
G	17 1%	75 5%	302 21%	275 19%	774 54%	1,443 100%	4.19
H	11 2%	45 7%	145 24%	118 19%	289 48%	608 100%	4.03
I	8 2%	26 7%	71 19%	82 22%	178 49%	365 100%	4.08
J	1 1%	13 10%	29 22%	34 26%	54 41%	131 100%	3.97

Table 6: Clarity: VVS1

Color	1 = Fair	2 = Good	3 = Very Good	4 = Premium	5 = Ideal	Total	Mean Quality Level
D	3 1%	13 5%	52 21%	40 16%	144 57%	252 100%	4.23
E	3 0%	43 7%	170 26%	105 16%	335 51%	656 100%	4.11
F	5 1%	35 5%	174 24%	80 11%	440 60%	734 100%	4.25
G	3 0%	41 4%	190 19%	171 17%	594 59%	999 100%	4.31
H	1 0%	31 5%	115 20%	112 19%	326 56%	585 100%	4.25
I	1 0%	22 6%	69 19%	84 24%	179 50%	355 100%	4.18
J	1 1%	1 1%	19 26%	24 32%	29 39%	74 100%	4.07

Table 7: Clarity: I1

Color	1 = Fair	2 = Good	3 = Very Good	4 = Premium	5 = Ideal	Total	Mean Quality Level
D	4 10%	8 19%	5 12%	12 29%	13 31%	42 100%	3.52
E	9 9%	23 23%	22 22%	30 29%	18 18%	102 100%	3.25
F	35 24%	19 13%	13 9%	34 24%	42 29%	143 100%	3.20
G	53 35%	19 13%	16 11%	46 31%	16 11%	150 100%	2.69
H	52 32%	14 9%	12 7%	46 28%	38 23%	162 100%	3.02
I	34 37%	9 10%	8 9%	24 26%	17 18%	92 100%	2.79
J	23 46%	4 8%	8 16%	13 26%	2 4%	50 100%	2.34

Table 8: Clarity: IF

Color	1 = Fair	2 = Good	3 = Very Good	4 = Premium	5 = Ideal	Total	Mean Quality Level
D	3 4%	9 12%	23 32%	10 14%	28 38%	73 100%	3.70
E	0 0%	9 6%	43 27%	27 17%	79 50%	158 100%	4.11
F	4 1%	15 4%	67 17%	31 8%	268 70%	385 100%	4.41
G	2 0%	22 3%	79 12%	87 13%	491 72%	681 100%	4.53
H	0 0%	4 1%	29 10%	40 13%	226 76%	299 100%	4.63
I	0 0%	6 4%	19 13%	23 16%	95 66%	143 100%	4.45
J	0 0%	6 12%	8 16%	12 24%	25 49%	51 100%	4.10