

# R Basic Notes.2

Charlene Garridos

2023-03-27

```
names <- c("Charl", "Garridos")
ages <- c(20, 20)
(df <- data.frame(names, ages))
```

```
##      names ages
## 1    Charl   20
## 2 Garridos   20
```

```
str(df)
```

```
## 'data.frame':   2 obs. of  2 variables:
## $ names: chr   "Charl" "Garridos"
## $ ages : num   20  20
```

```
dim(df)
```

```
## [1] 2 2
```

## url of data set

```
url <- "https://raw.githubusercontent.com/msuiitdmsgabriel/datasets-regression/main/salespeople.csv"
```

## load the data set and store it as a dataframe called salespeople

```
salespeople <- read.csv(url)
local_salespeople <- read.csv("salespeople.csv")
```

```
dim(salespeople)
```

```
## [1] 351  4
```

```
head(salespeople)
```

```
##      promoted sales customer_rate performance
## 1           0   594           3.94           2
## 2           0   446           4.06           3
## 3           1   674           3.83           4
## 4           0   525           3.62           2
## 5           1   657           4.40           3
## 6           1   918           4.54           2
```

```
salespeople$sales[6]
```

```
## [1] 918
```

```
salespeople$sales
```

```
##      [1] 594 446 674 525 657 918 318 364 342 387 527 716 557 450 344 372 258 338
##     [19] 410 937 702 469 535 342 819 736 330 274 341 717 478 487 239 825 400 728
##     [37] 773 425 943 510 389 270 945 497 329 389 475 383 432 619 578 411 445 440
##     [55] 359 419 840 393 754 441 803 444 753 688 431 511 464 473 532 280 342 320
##     [73] 531 373 547 611 825 431 401 517 803 586 444 693 659 416 423 756 245 419
##     [91] 757 617 909 516 317 425 528 416 645 390 393 394 387 450 487 607 369 489
##    [109] 324 417 694 651 395 442 422 404 381 501 944 753 591 735 538 451 477 436
##    [127] 738 902 464 944 285 453 382 414 335 935 203 348 800 436 360 674 425 901
##    [145] 453 350 362 486 471 459 506 262 825 291 464 802 818 736 364 308 862 349
##    [163] 375 423 938 456 517 373 898 777 470 545 699 697 300 677 497 669 596 492
##    [181] 346 590 592 780 432 418 662 678 716 330 414 416 403 362 284 363 655 597
##    [199] 794 818 409 681 606 489 475 590 396 420 857 371 421 828 594 533 462 392
##    [217] 475 752 659 650 496 211 898 388 383 455 319 756 377 940 757 469 394 484
##    [235] 491 547 519 739 479 943 742 357 432 584 595 401 460 753 466 362 361 338
##    [253] 882 293 922 793 787 400 516 295 307 151 441 406 270 680 662 347 453 309
##    [271] 592 540 886 420 718 284 323 513 841 362 842 321 516 428 383 521 358 489
##    [289] 252 720 610 871 594 522 379 454 450 317 835 297 516 355 858 305 410 707
##    [307] 798 265 576 448 590 456 930 412 286 440 546 385 544 505 732 506 394 674
##    [325] 458 251 429 348 789 795 509 754 580 289 390 787 241 522 412 359 489 940
##    [343] 592 796 653 459 586 401 500 373  NA
```

```
View(salespeople)
```

```
salespeople[34, 4]
```

```
## [1] 3
```

```
salespeople[34,]
```

```
##      promoted sales customer_rate performance
## 34           1   825           3.32           3
```

```
salespeople[,4]
```

```
## [1] 2 3 4 2 3 2 3 1 3 3 3 3 2 3 2 3 1 4 2 2 4 2 2 1 2
## [26] 4 2 1 2 2 2 1 4 3 2 3 3 1 4 3 4 2 4 3 3 4 3 2 3 3
## [51] 4 4 3 2 1 3 4 1 3 2 3 2 4 2 4 2 3 2 1 2 2 3 4 2 1
## [76] 4 2 3 2 3 3 1 4 3 1 3 3 4 2 2 3 1 3 1 1 3 2 1 2 4
## [101] 1 2 3 3 3 4 1 2 3 1 2 4 2 1 3 3 4 4 2 3 4 4 3 2 3
## [126] 2 3 4 1 4 3 2 2 2 3 3 2 3 2 3 1 3 3 3 2 3 2 1 2 3
## [151] 3 2 3 2 3 2 3 3 3 1 4 4 2 3 3 1 2 1 4 3 3 4 4 3 2
## [176] 3 1 4 2 3 2 2 3 4 2 2 4 3 3 1 1 2 1 3 3 1 3 3 3 1
## [201] 1 1 3 2 3 3 2 2 2 2 3 4 1 2 2 3 3 2 2 2 3 1 3 1 2
## [226] 3 4 3 3 3 3 3 1 2 4 2 4 3 2 4 2 1 3 2 2 3 2 2 2 2
## [251] 2 3 3 2 1 2 3 2 4 2 1 2 2 1 2 4 2 3 1 1 2 4 3 4 4
## [276] 2 3 3 4 1 3 1 3 4 3 1 2 3 2 3 4 2 3 2 3 2 2 2 1 4
## [301] 3 2 3 3 3 1 3 2 3 1 3 3 4 2 1 1 1 3 1 1 2 3 4 2 4
## [326] 2 1 3 3 1 3 4 1 3 3 1 2 2 2 2 3 4 4 3 3 3 2 3 2 1
## [351] NA
```

```
str(salespeople)
```

```
## 'data.frame': 351 obs. of 4 variables:
## $ promoted : int 0 0 1 0 1 1 0 0 0 0 ...
## $ sales : int 594 446 674 525 657 918 318 364 342 387 ...
## $ customer_rate: num 3.94 4.06 3.83 3.62 4.4 4.54 3.09 4.89 3.74 3 ...
## $ performance : int 2 3 4 2 3 2 3 1 3 3 ...
```

```
summary(salespeople)
```

```
## promoted sales customer_rate performance
## Min. :0.0000 Min. :151.0 Min. :1.000 Min. :1.0
## 1st Qu.:0.0000 1st Qu.:389.2 1st Qu.:3.000 1st Qu.:2.0
## Median :0.0000 Median :475.0 Median :3.620 Median :3.0
## Mean :0.3219 Mean :527.0 Mean :3.608 Mean :2.5
## 3rd Qu.:1.0000 3rd Qu.:667.2 3rd Qu.:4.290 3rd Qu.:3.0
## Max. :1.0000 Max. :945.0 Max. :5.000 Max. :4.0
## NA's :1 NA's :1 NA's :1
```

```
sum(is.na(salespeople))
```

```
## [1] 3
```

```
is.na(salespeople)
```

```
complete.cases(salespeople)
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [13] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [25] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [37] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
## [49] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [61] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [73] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [85] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [97] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [109] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [121] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [133] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [145] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [157] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [169] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [181] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [193] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [205] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [217] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [229] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [241] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [253] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [265] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [277] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [289] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [301] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [313] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [325] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [337] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [349] TRUE TRUE FALSE
```

```
salespeople[complete.cases(salespeople),]
```

```
salespeople <- salespeople[complete.cases(salespeople),]
```

```
is.na(salespeople)
```

```
sum(is.na(salespeople))
```

```
## [1] 0
```

```
unique(salespeople$performance)
```

```
## [1] 2 3 4 1
```

```
salespeople$performance
```

```
## [1] 2 3 4 2 3 2 3 1 3 3 3 3 2 3 2 3 1 4 2 2 4 2 2 1 2 4 2 1 2 2 2 1 4 3 2 3 3
## [38] 1 4 3 4 2 4 3 3 4 3 2 3 3 4 4 3 2 1 3 4 1 3 2 3 2 4 2 4 2 3 2 1 2 2 3 4 2
## [75] 1 4 2 3 2 3 3 1 4 3 1 3 3 4 2 2 3 1 3 1 1 3 2 1 2 4 1 2 3 3 3 4 1 2 3 1 2
## [112] 4 2 1 3 3 4 4 2 3 4 4 3 2 3 2 3 4 1 4 3 2 2 2 3 3 2 3 2 3 1 3 3 3 2 3 2 1
## [149] 2 3 3 2 3 2 3 2 3 3 3 1 4 4 2 3 3 1 2 1 4 3 3 4 4 3 2 3 1 4 2 3 2 2 3 4 2
## [186] 2 4 3 3 1 1 2 1 3 3 1 3 3 3 1 1 1 3 2 3 3 2 2 2 2 3 4 1 2 2 3 3 2 2 2 3 1
## [223] 3 1 2 3 4 3 3 3 3 3 1 2 4 2 4 3 2 4 2 1 3 2 2 3 2 2 2 2 2 3 3 2 1 2 3 2 4
## [260] 2 1 2 2 1 2 4 2 3 1 1 2 4 3 4 4 2 3 3 4 1 3 1 3 4 3 1 2 3 2 3 4 2 3 2 3 2
## [297] 2 2 1 4 3 2 3 3 3 1 3 2 3 1 3 3 4 2 1 1 1 3 1 1 2 3 4 2 4 2 1 3 3 1 3 4 1
## [334] 3 3 1 2 2 2 2 3 4 4 3 3 3 2 3 2 1
```

```
as.factor(salespeople$performance)
```

```
## [1] 2 3 4 2 3 2 3 1 3 3 3 3 2 3 2 3 1 4 2 2 4 2 2 1 2 4 2 1 2 2 2 1 4 3 2 3 3
## [38] 1 4 3 4 2 4 3 3 4 3 2 3 3 4 4 3 2 1 3 4 1 3 2 3 2 4 2 4 2 3 2 1 2 2 3 4 2
## [75] 1 4 2 3 2 3 3 1 4 3 1 3 3 4 2 2 3 1 3 1 1 3 2 1 2 4 1 2 3 3 3 4 1 2 3 1 2
## [112] 4 2 1 3 3 4 4 2 3 4 4 3 2 3 2 3 4 1 4 3 2 2 2 3 3 2 3 2 3 1 3 3 3 2 3 2 1
## [149] 2 3 3 2 3 2 3 2 3 3 3 1 4 4 2 3 3 1 2 1 4 3 3 4 4 3 2 3 1 4 2 3 2 2 3 4 2
## [186] 2 4 3 3 1 1 2 1 3 3 1 3 3 3 1 1 1 3 2 3 3 2 2 2 2 3 4 1 2 2 3 3 2 2 2 3 1
## [223] 3 1 2 3 4 3 3 3 3 3 1 2 4 2 4 3 2 4 2 1 3 2 2 3 2 2 2 2 2 3 3 2 1 2 3 2 4
## [260] 2 1 2 2 1 2 4 2 3 1 1 2 4 3 4 4 2 3 3 4 1 3 1 3 4 3 1 2 3 2 3 4 2 3 2 3 2
## [297] 2 2 1 4 3 2 3 3 3 1 3 2 3 1 3 3 4 2 1 1 1 3 1 1 2 3 4 2 4 2 1 3 3 1 3 4 1
## [334] 3 3 1 2 2 2 2 3 4 4 3 3 3 2 3 2 1
## Levels: 1 2 3 4
```

```
salespeople$performance <- as.factor(salespeople$performance)
str(salespeople)
```

```
## 'data.frame': 350 obs. of 4 variables:
## $ promoted : int 0 0 1 0 1 1 0 0 0 0 ...
## $ sales : int 594 446 674 525 657 918 318 364 342 387 ...
## $ customer_rate: num 3.94 4.06 3.83 3.62 4.4 4.54 3.09 4.89 3.74 3 ...
## $ performance : Factor w/ 4 levels "1","2","3","4": 2 3 4 2 3 2 3 1 3 3 ...
```

## Manipulating Dataframes

```
(sales_720 <- subset(salespeople, subset = sales == 720))
```

```
## promoted sales customer_rate performance
## 290 1 720 3.76 3
```

```
unique(salespeople$sales)
```

```
## [1] 594 446 674 525 657 918 318 364 342 387 527 716 557 450 344 372 258 338
## [19] 410 937 702 469 535 819 736 330 274 341 717 478 487 239 825 400 728 773
## [37] 425 943 510 389 270 945 497 329 475 383 432 619 578 411 445 440 359 419
## [55] 840 393 754 441 803 444 753 688 431 511 464 473 532 280 320 531 373 547
## [73] 611 401 517 586 693 659 416 423 756 245 757 617 909 516 317 528 645 390
## [91] 394 607 369 489 324 417 694 651 395 442 422 404 381 501 944 591 735 538
## [109] 451 477 436 738 902 285 453 382 414 335 935 203 348 800 360 901 350 362
## [127] 486 471 459 506 262 291 802 818 308 862 349 375 938 456 898 777 470 545
## [145] 699 697 300 677 669 596 492 346 590 592 780 418 662 678 403 284 363 655
## [163] 597 794 409 681 606 396 420 857 371 421 828 533 462 392 752 650 496 211
## [181] 388 455 319 377 940 484 491 519 739 479 742 357 584 595 460 466 361 882
## [199] 293 922 793 787 295 307 151 406 680 347 309 540 886 718 323 513 841 842
## [217] 321 428 521 358 252 720 610 871 522 379 454 835 297 355 858 305 707 798
## [235] 265 576 448 930 412 286 546 385 544 505 732 458 251 429 789 795 509 580
## [253] 289 241 796 653 500
```

```
(subset(salespeople, subset = sales >= 700))
```

##	promoted	sales	customer_rate	performance
## 6	1	918	4.54	2
## 12	1	716	3.16	3
## 20	1	937	5.00	2
## 21	1	702	3.53	4
## 25	1	819	4.45	2
## 26	1	736	3.94	4
## 30	1	717	2.98	2
## 34	1	825	3.32	3
## 36	1	728	2.66	3
## 37	1	773	4.89	3
## 39	1	943	4.40	4
## 43	1	945	4.31	4
## 57	1	840	5.00	4
## 59	1	754	3.74	3
## 61	1	803	4.89	3
## 63	1	753	5.00	4
## 77	1	825	4.66	2
## 81	1	803	4.15	3
## 88	1	756	3.55	4
## 91	1	757	3.11	3
## 93	1	909	3.21	3
## 119	1	944	5.00	2
## 120	1	753	4.43	3
## 122	1	735	4.03	4
## 127	1	738	3.05	3
## 128	1	902	5.00	4
## 130	1	944	3.92	4
## 136	1	935	5.00	3
## 139	1	800	4.24	2
## 144	1	901	2.70	3
## 153	1	825	4.95	3
## 156	1	802	3.78	2
## 157	1	818	4.24	3
## 158	1	736	3.78	3
## 161	1	862	4.17	4
## 165	1	938	3.69	3
## 169	1	898	2.26	4
## 170	1	777	4.86	3
## 184	1	780	4.15	4
## 189	1	716	3.44	3
## 199	1	794	3.83	3
## 200	1	818	2.70	1
## 209	1	857	3.85	2
## 212	1	828	1.37	4
## 218	1	752	4.89	2
## 223	1	898	3.51	3
## 228	1	756	3.09	3
## 230	1	940	2.82	3
## 231	1	757	3.55	3
## 238	1	739	3.99	3

##	240	1	943	3.21	4
##	241	1	742	4.17	2
##	248	1	753	4.89	2
##	253	1	882	2.63	3
##	255	1	922	4.15	1
##	256	1	793	4.08	2
##	257	1	787	2.56	3
##	273	1	886	4.68	3
##	275	1	718	4.03	4
##	279	1	841	5.00	4
##	281	1	842	3.99	3
##	290	1	720	3.76	3
##	292	1	871	5.00	2
##	299	1	835	3.90	1
##	303	1	858	3.67	3
##	306	1	707	2.38	1
##	307	1	798	4.72	3
##	313	1	930	4.22	4
##	321	1	732	3.57	2
##	329	1	789	3.71	3
##	330	1	795	4.31	1
##	332	1	754	4.33	4
##	336	1	787	3.14	1
##	342	1	940	5.00	4
##	344	1	796	5.00	3

```
high_sales <- subset(salespeople, subset = sales >= 700)
head(high_sales)
```

##	promoted	sales	customer_rate	performance	
##	6	1	918	4.54	2
##	12	1	716	3.16	3
##	20	1	937	5.00	2
##	21	1	702	3.53	4
##	25	1	819	4.45	2
##	26	1	736	3.94	4