

tidy data

Yunting Chiu

11/1/2020

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2    v purrr  0.3.4
## v tibble  3.0.3    v dplyr  1.0.2
## v tidyr   1.1.2    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
ws <- read.csv(file = "../data/Datafiniti_Womens_Shoes_Jun19.csv")
```

```
ws %>%
```

```
  select(prices.size) %>%
  distinct()
```

```
##           prices.size
## 1                    8
## 2                 6 M US
## 3                    5
## 4                 9.5 BM US
## 5                    11
## 6                  11 M
## 7                    9
## 8                  Size 6
## 9                  Size 5
## 10                 Size 8
## 11                   7
## 12                 6 BM US
## 13                   7M
## 14                  5.5
## 15                   6
## 16                 8 BM US
## 17                   12
## 18                 9 BM US
## 19                   6 M
## 20                  7.5
## 21                  US 6.5
## 22                   10
## 23                   6.5
## 24                 10 BM US
## 25                  9.5
```

## 26	8M
## 27	7.5 R
## 28	8 R
## 29	8.5 BM US
## 30	5.5 BM US
## 31	US 10.5W
## 32	US 7.5
## 33	9M
## 34	US 8
## 35	13
## 36	8.5
## 37	9.5 N
## 38	6M
## 39	7.5M
## 40	6.5 B US
## 41	8.5 M US
## 42	8 M
## 43	6.5M
## 44	4
## 45	10.5M
## 46	Size 9
## 47	US 8.5W
## 48	10.5
## 49	7 BM US
## 50	5.5M
## 51	6.5 BM US Women
## 52	5W
## 53	9.5 Medium BM
## 54	8W
## 55	7 C/D US
## 56	9 Medium BM
## 57	10 BM US Women
## 58	14M
## 59	Size 7
## 60	11M
## 61	10M
## 62	US 7
## 63	US 8.5
## 64	7.5 M
## 65	7.5 Medium BM
## 66	Size 10
## 67	6 W
## 68	9 W
## 69	8 M US
## 70	10 Medium BM
## 71	5M
## 72	One Size
## 73	7 S
## 74	10 N
## 75	S
## 76	M
## 77	40
## 78	9.5M
## 79	6.5 Medium BM

## 80	7.5 WW
## 81	10.5 N
## 82	6.5 WW
## 83	10.5 W
## 84	16
## 85	6.5 W
## 86	9 M
## 87	9.5 M
## 88	6.5 M
## 89	8.5 M
## 90	8.5 S
## 91	US 5.5
## 92	Size 12
## 93	8.5M
## 94	11.5
## 95	3
## 96	6 Medium BM
## 97	7 M US
## 98	11 Medium BM
## 99	7 Medium BM
## 100	7.5 M US
## 101	6.5 M US
## 102	US 9
## 103	US 10
## 104	9 M US
## 105	10.0M
## 106	41
## 107	9.5W
## 108	10.5 B US
## 109	600
## 110	6.5 BM US
## 111	7.5 US / 37.5 EU
## 112	11N
## 113	39
## 114	42
## 115	37
## 116	14
## 117	9 WW
## 118	10.5 M
## 119	7.5 N
## 120	5.5 M
## 121	10 WW
## 122	7 W
## 123	Size 6.5
## 124	5.5 M US
## 125	5 Medium BM
## 126	7.5 BM US
## 127	10 M
## 128	7.5 S
## 129	12 M
## 130	9.5 W
## 131	8 Standard
## 132	11 BM US
## 133	6 2W US

## 134	10.5 M US
## 135	7.5 Standard
## 136	35-36
## 137	8 Medium BM
## 138	9 M US Women
## 139	36 EU
## 140	38
## 141	8.5 Medium BM
## 142	36 EU / 5.5-6 US
## 143	36
## 144	10 Boot
## 145	Kids 5 Plus
## 146	Kids 6 Plus
## 147	7 M
## 148	36 M EU / 6 BM US
## 149	9 Wide CDW
## 150	Purple 10
## 151	8 M US Women
## 152	8.5 M US Women
## 153	10 M US Women
## 154	7.5 M US Women
## 155	7 M US Women
## 156	6 M US Women
## 157	12 WW US
## 158	6.5 M US Women
## 159	11 M US
## 160	US 4
## 161	US 5
## 162	5 BM US
## 163	8 BM
## 164	11 W
## 165	10 M US
## 166	US 11
## 167	11 US / 41 EU
## 168	2XL
## 169	8 DM US
## 170	9 M US Toddler
## 171	2 M US Little Kid
## 172	9 W US
## 173	Womenfts Shoe 9.5 AAN
## 174	Womenfts US Shoe Size 9
## 175	X-Large
## 176	Size 14
## 177	Size 9.5
## 178	Size 11
## 179	8 2W US
## 180	6 US
## 181	40 EU / 9 US
## 182	9.5 M US
## 183	4M
## 184	8 W
## 185	10 W
## 186	US 6
## 187	11.5 M

## 188	11.5 W
## 189	7 N
## 190	Size 7.5
## 191	Womenfts US Shoe Size 12
## 192	Womenfts US Shoe Size 8
## 193	6 Medium B M US
## 194	17
## 195	40 EU / 9-9.5 US
## 196	6.5 D
## 197	13 WW
## 198	5 M
## 199	8.5 N
## 200	5 W
## 201	Child Medium
## 202	9 N
## 203	44
## 204	38 Regular EU
## 205	35
## 206	35 5 M
## 207	43
## 208	4.5
## 209	11.5 S
## 210	10 N US
## 211	3.5
## 212	38 M EU / 7.5-8 BM US
## 213	41 M EU / 9.5-10 BM US
## 214	6 Boot
## 215	2 UK US
## 216	10W
## 217	6N
## 218	5 Medium B M US
## 219	12M
## 220	7W
## 221	9 B US
## 222	8 B US
## 223	6 BM US Women
## 224	7.5 BM US Women
## 225	7 BM US Women
## 226	8.5 BM US Women
## 227	7 WW
## 228	Extra Large
## 229	Medium
## 230	15
## 231	7.5W
## 232	US 9.5
## 233	6.5 Narrow AAN
## 234	7 B
## 235	11 Standard
## 236	40 M EU / 10 BM US
## 237	37 M EU / 7-7.5 BM US
## 238	9 Narrow AAN
## 239	7.5 US
## 240	7 Wide CDW
## 241	10 M US Toddler

## 242	1 M US Little Kid
## 243	12 M US Little Kid
## 244	3 M US Little Kid
## 245	4.5M
## 246	Kids 8 Plus
## 247	11 B US
## 248	9 1/2 0950
## 249	6 B US
## 250	6 D US
## 251	7 2E US
## 252	5.5 M US Women
## 253	11.5WW
## 254	13WW
## 255	7.5 DM US
## 256	7 DM US
## 257	9.5 DM US
## 258	11 1/2 1150
## 259	11W
## 260	07.5 UK
## 261	9 WW US
## 262	6.5W
## 263	6W
## 264	9W
## 265	10.5W
## 266	8 N
## 267	7.5 B
## 268	8.5W
## 269	11.5M
## 270	10 B US
## 271	6.5 US / 37 EU
## 272	8.5 B M US
## 273	6 B M US
## 274	7 B M US
## 275	10 B M US
## 276	9.5-10
## 277	8.5-9
## 278	6 BM US Women / 4.5 DM US Men
## 279	Size 13
## 280	9.5 XX-Wide
## 281	7.0M
## 282	4.0M
## 283	10.5 S
## 284	6M Regular
## 285	37 EU / 6.5-7 US
## 286	10 Standard
## 287	5.5 Standard
## 288	5 Standard
## 289	6.5 Standard
## 290	5.5 Medium BM
## 291	6.5 US
## 292	7 US / 37.5 EU
## 293	6.5 W US
## 294	10.5 BM US
## 295	5-5.5

## 296	8.5 B US
## 297	11 Narrow AAN
## 298	6 Extra Wide E WW
## 299	Xlarge 9.5-10.5
## 300	36 M EU / 5.5-6 BM US
## 301	39 M EU / 8.5-9 BM US
## 302	41 EU
## 303	40 M EU / 9 BM US
## 304	L
## 305	Small/Medium
## 306	12 W
## 307	Large 9
## 308	6.5 Boot
## 309	11 Boot
## 310	10.5 Boot
## 311	9.5 Boot
## 312	8.5 Boot
## 313	ML
## 314	7.5N
## 315	XS
## 316	12W
## 317	5.5W
## 318	US 7W
## 319	11 WW
## 320	5.5 WW
## 321	12 N
## 322	12 S
## 323	4.5 M
## 324	6 S
## 325	11 S
## 326	9.5 S
## 327	6.5 N
## 328	5.5N
## 329	37 M
## 330	6 Standard
## 331	5.0M
## 332	11 N
## 333	10 3W
## 334	35 EU / 5-5.5 US
## 335	42 EU / 10.5-11 US
## 336	37 M EU / 6.5-7 BM US
## 337	41 M EU / 10.5-11 BM US
## 338	38 M EU
## 339	38 EU Women
## 340	40 M EU / 9-9.5 BM US
## 341	9.5 WW
## 342	8.5 WW
## 343	6 WW
## 344	7.5 W
## 345	8.5 W
## 346	8 WW
## 347	5 US
## 348	10.5 WW
## 349	39 M EU Women

## 350	37 M EU Women
## 351	38 M EU Women
## 352	11.5 N
## 353	Size
## 354	8.5 US / 38.5 EU
## 355	55 M US Women
## 356	9.5 Wide CDW
## 357	9.5 N US
## 358	9 S
## 359	7 3W
## 360	4 M
## 361	9 US / 39 EU
## 362	13 W
## 363	9.5 BM
## 364	10 BM
## 365	11 N US
## 366	5 M US
## 367	9.5 B US
## 368	Men 12/Women 13.5
## 369	12 WW
## 370	8 S
## 371	9 B M US
## 372	12 XX-Wide4E US
## 373	9 XX-Wide4E US
## 374	10 S
## 375	10 Narrow AAN
## 376	Uk 3 / US 5
## 377	11 1100
## 378	6 1/2 0650
## 379	6 0600
## 380	6 2E US
## 381	Kids 7 Plus
## 382	6.5 S
## 383	EU36US5.5-6
## 384	39 EU
## 385	6 N
## 386	38 EU / 7.5-8 US
## 387	Turquoise
## 388	M 065 Medium US
## 389	11.5 WW
## 390	40 EU
## 391	37 Medium BM
## 392	37 EU
## 393	40 MediumBM
## 394	36 Medium BM
## 395	5.5 3W
## 396	7 Women US
## 397	Womenfts Shoe 12.5 DW
## 398	5 WW
## 399	7 B US
## 400	7 0700
## 401	5.5 W
## 402	US Size 5
## 403	8 US

## 404	8 B
## 405	13M
## 406	11 Wide CDW
## 407	12 Narrow AAN
## 408	8.5 X-Wide
## 409	9 DM US
## 410	8.5 US / 39 EU
## 411	Size 4
## 412	13-14
## 413	XL
## 414	6 Medium D M
## 415	9 Medium D M
## 416	9.5 Extra Wide E WW
## 417	7 Extra Wide E WW
## 418	7.5 Wide CDW
## 419	7 US
## 420	13 N
## 421	13 S
## 422	38 M EU / 8-8.5 BM US
## 423	8.5 2W US
## 424	7 Standard
## 425	10 US
## 426	6 N US
## 427	5.5 US
## 428	9 Wide
## 429	11 US
## 430	8 Wide
## 431	9 US
## 432	8.5 W US
## 433	7.5 C/D US
## 434	8 C/D US
## 435	9 C/D US
## 436	Womenfts Shoe 10.5
## 437	8 XX-Wide
## 438	10 4E
## 439	9 XX-Wide
## 440	11 X-Wide
## 441	7.5 XX-Wide
## 442	9.5 X-Wide
## 443	9 X-Wide
## 444	11 XX-Wide
## 445	8.5 XX-Wide
## 446	7 XX-Wide
## 447	7 X-Wide
## 448	6.5 X-Wide
## 449	6 XX-Wide
## 450	7.5 X-Wide
## 451	10.5 XX-Wide
## 452	6.5 XX-Wide
## 453	10.5 X-Wide
## 454	10 XX-Wide
## 455	8 X-Wide
## 456	10 B
## 457	Womenfts Shoe 11

```
## 458          8.5 Standard
## 459      Womenfts Shoe 10  AAN
## 460          12 Medium BM
## 461          Size XS
## 462          10M / 8.5M
## 463      39 M EU / 8.5 BM US
## 464          LG
## 465      Womenfts Shoe 10
## 466          6 X-Wide
## 467          9.5-2E
```

```
ws %>%
  # select(prices.size) %>%
  mutate(sizesUS = str_extract(prices.size, "\\d{1,}")) %>%
  filter(!is.na(sizesUS)) -> ws
ws %>%
  distinct(sizesUS)
```

```
##      sizesUS
## 1          8
## 2          6
## 3          5
## 4          9
## 5         11
## 6          7
## 7         12
## 8         10
## 9         13
## 10         4
## 11         14
## 12        40
## 13        16
## 14         3
## 15        41
## 16       600
## 17        39
## 18        42
## 19        37
## 20        35
## 21        36
## 22        38
## 23         2
## 24        17
## 25        44
## 26        43
## 27        15
## 28         1
## 29        07
## 30        55
## 31       065
```

```
# convert Europe sizes to the US sizes
ws %>%
  mutate(sizesUS = recode(sizesUS, "44" = "12",
                           "43" = "11",
```

```

      "42" = "11",
      "41" = "10",
      "40" = "9",
      "39" = "8",
      "38" = "7.5",
      "37" = "6.5",
      "36" = "6",
      "35" = "5"
    ),
    sizesUS = as.numeric(sizesUS)) -> ws

```

```

# testing
ws %>%
  arrange(desc(sizesUS)) %>%
  distinct(sizesUS)

```

```

##      sizesUS
## 1      600.0
## 2       65.0
## 3       55.0
## 4       17.0
## 5       16.0
## 6       15.0
## 7       14.0
## 8       13.0
## 9       12.0
## 10      11.0
## 11      10.0
## 12       9.0
## 13       8.0
## 14       7.5
## 15       7.0
## 16       6.5
## 17       6.0
## 18       5.0
## 19       4.0
## 20       3.0
## 21       2.0
## 22       1.0

```

```

ws %>%
  mutate(free.shipping = ifelse(str_length(prices.shipping) > 0, "NotFree", "Free"))%>%
  separate(col = prices.offer, into = c("channel", "discount"), sep = ":")%>%
  mutate(discount = ifelse(is.na(discount), 0, as.numeric(substr(discount, 1, str_length(discount)-5))))/
  mutate(channel = ifelse(substr(channel, 1, 3) == "Was", "In store only", channel)) -> ws

```

```

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 7166 rows [1, 2,
## 3, 6, 7, 10, 12, 13, 14, 16, 17, 18, 19, 22, 24, 25, 26, 27, 29, 30, ...].

```