HW05: Factor and Figure Management

Carleena Ortega

14/10/2019

Contents


```
suppressPackageStartupMessages(library(gapminder))
suppressPackageStartupMessages(library(tibble))
suppressPackageStartupMessages(library(forcats))
suppressPackageStartupMessages(library(ggplot2))
suppressPackageStartupMessages(library(tidyverse))
suppressPackageStartupMessages(library(dplyr))
suppressPackageStartupMessages(library(knitr))
suppressPackageStartupMessages(library(kableExtra))
```

Exercise 1: Explain the value of the here::here package

In your own words, summarize the value of the here::here package in 250 words or fewer.

Exercise 2: Factor management

We choose the `gapminder` dataset and select `continent` as the factor to explore. The code below determines whether gapminder's continent is a factor (TRUE) or not (FALSE).

```
is.factor(gapminder$continent)
```

```
## [1] TRUE
```

Now that we have verified that `continent` is truly a factor, our next step is to explore the effects of re-leveling a factor with the gapminder data set.

We drop the observations from the continent of Oceania.

Drop Oceania. Filter the Gapminder data to remove observations associated with the continent of Oceania. /Additionally, remove unused factor levels. Provide concrete information on the data before and after removing these rows and Oceania; address the number of rows and the levels of the affected factors.

Reorder the levels of country or continent. Use the forcats package to change the order of the factor levels, based on summarized information of one of the quantitative variables. Consider experimenting with

a summary statistic beyond the most basic choice of the mean/median. Use the forcats package in the tidyverse for this, rather than the baseR function as.factor.

Explore the effects of re-leveling a factor in a tibble by:

comparing the results of arrange on the original and re-leveled factor. Plotting a figure of before/after re-leveling the factor (make sure to assign the factor to an aesthetic of your choosing). These explorations should involve the data, the factor levels, and at least two figures (before and after.

```
gapminder %>%
  filter(continent !="Oceania") %>%
  arrange(continent) %>%
knitr::kable() %>%
kable_styling("striped")
```

country	continent	year	lifeExp	pop	gdpPercap
Algeria	Africa	1952	43.07700	9279525	2449.0082
Algeria	Africa	1957	45.68500	10270856	3013.9760
Algeria	Africa	1962	48.30300	11000948	2550.8169
Algeria	Africa	1967	51.40700	12760499	3246.9918
Algeria	Africa	1972	54.51800	14760787	4182.6638
Algeria	Africa	1977	58.01400	17152804	4910.4168
Algeria	Africa	1982	61.36800	20033753	5745.1602
Algeria	Africa	1987	65.79900	23254956	5681.3585
Algeria	Africa	1992	67.74400	26298373	5023.2166
Algeria	Africa	1997	69.15200	29072015	4797.2951
Algeria	Africa	2002	70.99400	31287142	5288.0404
Algeria	Africa	2007	72.30100	33333216	6223.3675
Angola	Africa	1952	30.01500	4232095	3520.6103
Angola	Africa	1957	31.99900	4561361	3827.9405
Angola	Africa	1962	34.00000	4826015	4269.2767
Angola	Africa	1967	35.98500	5247469	5522.7764
Angola	Africa	1972	37.92800	5894858	5473.2880
Angola	Africa	1977	39.48300	6162675	3008.6474
Angola	Africa	1982	39.94200	7016384	2756.9537
Angola	Africa	1987	39.90600	7874230	2430.2083
Angola	Africa	1992	40.64700	8735988	2627.8457
Angola	Africa	1997	40.96300	9875024	2277.1409
Angola	Africa	2002	41.00300	10866106	2773.2873
Angola	Africa	2007	42.73100	12420476	4797.2313
Benin	Africa	1952	38.22300	1738315	1062.7522
Benin	Africa	1957	40.35800	1925173	959.6011
Benin	Africa	1962	42.61800	2151895	949.4991
Benin	Africa	1967	44.88500	2427334	1035.8314
Benin	Africa	1972	47.01400	2761407	1085.7969
Benin	Africa	1977	49.19000	3168267	1029.1613
Benin	Africa	1982	50.90400	3641603	1277.8976
Benin	Africa	1987	52.33700	4243788	1225.8560
Benin	Africa	1992	53.91900	4981671	1191.2077
Benin	Africa	1997	54.77700	6066080	1232.9753
Benin	Africa	2002	54.40600	7026113	1372.8779
Benin	Africa	2007	56.72800	8078314	1441.2849
Botswana	Africa	1952	47.62200	442308	851.2411
Botswana	Africa	1957	49.61800	474639	918.2325
Botswana	Africa	1962	51.52000	512764	983.6540
Botswana	Africa	1967	53.29800	553541	1214.7093
Botswana	Africa	1972	56.02400	619351	2263.6111
Botswana	Africa	1977	59.31900	781472	3214.8578
Botswana	Africa	1982	61.48400	970347	4551.1421
Botswana	Africa	1987	63.62200	1151184	6205.8839
Botswana	Africa	1992	62.74500	1342614	7954.1116
Botswana	Africa	1997	52.55600	1536536	8647.1423
Botswana	Africa	2002	46.63400	1630347	11003.6051
Botswana	Africa	2007	50.72800	1639131	12569.8518
Burkina Faso	Africa	1952	31.97500	4469979	543.2552
Burkina Faso	Africa	1957	34.90600	4713416	617.1835
Burkina Faso	Africa	1962	37.81400	4919632	722.5120
Burkina Faso	Africa	1967	40.69700	5127935	794.8266
Burkina Faso	Africa	1972	43.59100	5433886	854.7360
Burkina Faso	Africa	1977	46.13700	5889574	743.3870
Burkina Faso	Africa	1982	48.12200	6634596	807.1986
Burkina Faso	Africa	1987	49.55700	7586551	912.0631
Burkina Faso	Africa	1992	50.26000	8878303	931.7528

```r
b<-gapminder %>%
  filter(continent !="Oceania") %>%
  droplevels("country")
knitr::kable(b) %>%
  kable_styling("striped")
```

country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1952	28.80100	8425333	779.4453
Afghanistan	Asia	1957	30.33200	9240934	820.8530
Afghanistan	Asia	1962	31.99700	10267083	853.1007
Afghanistan	Asia	1967	34.02000	11537966	836.1971
Afghanistan	Asia	1972	36.08800	13079460	739.9811
Afghanistan	Asia	1977	38.43800	14880372	786.1134
Afghanistan	Asia	1982	39.85400	12881816	978.0114
Afghanistan	Asia	1987	40.82200	13867957	852.3959
Afghanistan	Asia	1992	41.67400	16317921	649.3414
Afghanistan	Asia	1997	41.76300	22227415	635.3414
Afghanistan	Asia	2002	42.12900	25268405	726.7341
Afghanistan	Asia	2007	43.82800	31889923	974.5803
Albania	Europe	1952	55.23000	1282697	1601.0561
Albania	Europe	1957	59.28000	1476505	1942.2842
Albania	Europe	1962	64.82000	1728137	2312.8890
Albania	Europe	1967	66.22000	1984060	2760.1969
Albania	Europe	1972	67.69000	2263554	3313.4222
Albania	Europe	1977	68.93000	2509048	3533.0039
Albania	Europe	1982	70.42000	2780097	3630.8807
Albania	Europe	1987	72.00000	3075321	3738.9327
Albania	Europe	1992	71.58100	3326498	2497.4379
Albania	Europe	1997	72.95000	3428038	3193.0546
Albania	Europe	2002	75.65100	3508512	4604.2117
Albania	Europe	2007	76.42300	3600523	5937.0295
Algeria	Africa	1952	43.07700	9279525	2449.0082
Algeria	Africa	1957	45.68500	10270856	3013.9760
Algeria	Africa	1962	48.30300	11000948	2550.8169
Algeria	Africa	1967	51.40700	12760499	3246.9918
Algeria	Africa	1972	54.51800	14760787	4182.6638
Algeria	Africa	1977	58.01400	17152804	4910.4168
Algeria	Africa	1982	61.36800	20033753	5745.1602
Algeria	Africa	1987	65.79900	23254956	5681.3585
Algeria	Africa	1992	67.74400	26298373	5023.2166
Algeria	Africa	1997	69.15200	29072015	4797.2951
Algeria	Africa	2002	70.99400	31287142	5288.0404
Algeria	Africa	2007	72.30100	33333216	6223.3675
Angola	Africa	1952	30.01500	4232095	3520.6103
Angola	Africa	1957	31.99900	4561361	3827.9405
Angola	Africa	1962	34.00000	4826015	4269.2767
Angola	Africa	1967	35.98500	5247469	5522.7764
Angola	Africa	1972	37.92800	5894858	5473.2880
Angola	Africa	1977	39.48300	6162675	3008.6474
Angola	Africa	1982	39.94200	7016384	2756.9537
Angola	Africa	1987	39.90600	7874230	2430.2083
Angola	Africa	1992	40.64700	8735988	2627.8457
Angola	Africa	1997	40.96300	9875024	2277.1409
Angola	Africa	2002	41.00300	10866106	2773.2873
Angola	Africa	2007	42.73100	12420476	4797.2313
Argentina	Americas	1952	62.48500	17876956	5911.3151
Argentina	Americas	1957	64.39900	19610538	6856.8562
Argentina	Americas	1962	65.14200	21283783	7133.1660
Argentina	Americas	1967	65.63400	22934225	8052.9530
Argentina	Americas	1972	67.06500	24779799	9443.0385
Argentina	Americas	1977	68.48100	26983828	10079.0267
Argentina	Americas	1982	69.94200	29341374	8997.8974
Argentina	Americas	1987	70.77400	31620918	9139.6714
Argentina	Americas	1992	71.86800	33958947	9308.4187

```
gapminder
```

```
## # A tibble: 1,704 x 6
##    country     continent  year lifeExp      pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
##  1 Afghanistan Asia       1952    28.8  8425333      779.
##  2 Afghanistan Asia       1957    30.3  9240934      821.
##  3 Afghanistan Asia       1962    32.0 10267083      853.
##  4 Afghanistan Asia       1967    34.0 11537966      836.
##  5 Afghanistan Asia       1972    36.1 13079460      740.
##  6 Afghanistan Asia       1977    38.4 14880372      786.
##  7 Afghanistan Asia       1982    39.9 12881816      978.
##  8 Afghanistan Asia       1987    40.8 13867957      852.
##  9 Afghanistan Asia       1992    41.7 16317921      649.
## 10 Afghanistan Asia       1997    41.8 22227415      635.
## # ... with 1,694 more rows
```

Exercise 3: File input/output (I/O)

Task: Experiment with at least one of:

write_csv()/read_csv() (and/or TSV friends), saveRDS()/readRDS(), dput()/dget(). You are expected to create something new, probably by filtering or grouped-summarization of your dataset (for e.g., Singer, Gapminder, or another dataset), export it to disk and then reload it back in using one of the packages above. You should use here::here() for reading in and writing out.

With the imported data, play around with factor levels and use factors to order your data with one of your factors (i.e. non-alphabetically). For the I/O method(s) you chose, comment on whether or not your newly created file survived the round trip of writing to file then reading back in.

Exercise 4: Visualization design (20%)

Rubric:

5%: Writing 5%: Accuracy 10%: Vis quality Go back through your previous assignments and class participation activities and find figures you created prior to the last week of the course. Recreate at least one figure in light of something you learned in the recent class meetings about visualization design and color.

Task: Create a side-by-side plot and juxtapose your first attempt (show the original figure as-is) with a revised attempt after some time spent working on it and implementing principles of effective plotting principles. Comment and reflect on the differences.

Exercise 5: Writing figures to file (10%) 5%: Accuracy 5%: Code quality Task: Use ggsave() to explicitly save a plot to file. Include the exported plot as part of your repository and assignment.

Then, use [Alt text] (/path/to/img.png) to load and embed that file into your report. You can play around with various options, such as:

Arguments of ggsave(), such as width, height, resolution or text scaling. Various graphics devices, e.g. a vector vs. raster format. Explicit provision of the plot object p via ggsave(..., plot = p). Show a situation in which this actually matters.