

Address Missing Data

Step 1: Load your data.

```
library(tidyverse)
langDiv <- read.csv("language_diversity.csv")
gdp <- read.csv("country_gdp_data.csv", header = TRUE)
# perform a full join on the langDiv and gdp data sets
joinDF <- full_join(x = langDiv, y = gdp, by = "Country")

head(joinDF)
```

##	Country	Langs	Area	Population	Continent	GS	GDP
## 1	Algeria	18	2381741	25660	Africa	6.60	1452
## 2	Angola	42	1246700	10303	Africa	6.22	466
## 3	Australia	234	7713364	17336	Oceania	6.00	21635
## 4	Bangladesh	37	143998	118745	Asia	7.40	319
## 5	Benin	52	112622	4889	Africa	7.14	397
## 6	Bolivia	38	1098581	7612	Americas	6.92	887

Step 2: Identify complete rows.

Use the `complete.cases()` function to determine which rows of your data set are complete, which means that they don't contain NAs. The argument to `complete.cases` is the data set you want to examine:

```
temp <- complete.cases(joinDF)
head(temp)
```

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

Step 3: Keep rows that are complete.

Using `complete.cases()` in conjunction with `filter()` will allow you to keep only the rows that have data in every cell and are therefore complete:

```
joinDF %>%
  filter(complete.cases(joinDF))
```

##	Country	Langs	Area	Population	Continent	GS	GDP
## 1	Algeria	18	2381741	25660	Africa	6.60	1452
## 2	Angola	42	1246700	10303	Africa	6.22	466
## 3	Australia	234	7713364	17336	Oceania	6.00	21635
## 4	Bangladesh	37	143998	118745	Asia	7.40	319
## 5	Benin	52	112622	4889	Africa	7.14	397
## 6	Bolivia	38	1098581	7612	Americas	6.92	887
## 7	Botswana	27	581730	1348	Africa	4.60	3015
## 8	Brazil	209	8511965	153322	Americas	9.71	4794
## 9	Burkina Faso	75	274000	9242	Africa	5.17	238
## 10	CAR	94	622984	3127	Africa	8.08	348
## 11	Cambodia	18	181035	8442	Asia	8.44	311

## 12	Cameroon	275	475422	12239	Africa	9.17	716
## 13	Chad	126	1284000	5819	Africa	4.00	235
## 14	Colombia	79	1138914	33613	Americas	11.37	2946
## 15	Congo	60	342000	2346	Africa	9.60	789
## 16	Costa Rica	10	51100	3064	Americas	8.92	3278
## 17	Cote d'Ivoire	75	322463	12464	Africa	8.67	764
## 18	Cuba	1	110861	10736	Americas	7.46	2790
## 19	Ecuador	22	283561	10851	Americas	8.14	2135
## 20	Egypt	11	1001449	54688	Africa	0.89	1032
## 21	Ethiopia	112	1221900	53383	Africa	7.28	132
## 22	Gabon	40	267667	1212	Africa	8.79	5081
## 23	Ghana	73	238553	15509	Africa	8.79	818
## 24	Guatemala	52	108889	9467	Americas	9.31	1255
## 25	Guinea	29	245857	5931	Africa	7.38	668
## 26	Guyana	14	214969	800	Americas	12.00	1302
## 27	Honduras	9	112088	5265	Americas	8.54	828
## 28	India	405	3287590	849638	Asia	5.32	378
## 29	Indonesia	701	1904569	187765	Asia	10.67	1195
## 30	Kenya	58	580367	25905	Africa	7.26	491
## 31	Laos	93	236800	4262	Asia	7.14	382
## 32	Liberia	34	111369	2705	Africa	10.62	101
## 33	Libya	13	1759540	4712	Africa	2.43	5717
## 34	Madagascar	4	587041	11493	Africa	7.33	277
## 35	Malawi	14	118484	8556	Africa	5.80	250
## 36	Malaysia	140	329749	18333	Asia	11.92	4334
## 37	Mali	31	1240192	9507	Africa	3.59	295
## 38	Mauritania	8	1025520	2036	Africa	0.75	722
## 39	Mexico	243	1958201	87836	Americas	5.84	3829
## 40	Mozambique	36	801590	16084	Africa	6.07	163
## 41	Myanmar	105	676578	42561	Asia	6.93	180
## 42	Namibia	21	824292	1837	Africa	2.50	2382
## 43	Nepal	102	140797	19605	Asia	6.39	212
## 44	Nicaragua	7	130000	3999	Americas	8.13	893
## 45	Niger	21	1267000	7984	Africa	2.40	188
## 46	Nigeria	427	923768	112163	Africa	7.00	454
## 47	Oman	8	212457	1559	Asia	0.00	6193
## 48	Panama	13	75517	2466	Americas	9.20	3354
## 49	Papua New Guinea	862	462840	3772	Oceania	10.88	1486
## 50	Paraguay	21	406752	4397	Americas	10.25	1694
## 51	Peru	91	1285216	21998	Americas	2.65	2220
## 52	Philippines	168	300000	62868	Asia	10.34	1176
## 53	Saudi Arabia	8	2149690	14691	Asia	0.40	7641
## 54	Senegal	42	196722	7533	Africa	3.58	721
## 55	Sierra Leone	23	71740	4260	Africa	8.22	276
## 56	Solomon Islands	66	28896	3301	Oceania	12.00	1015
## 57	Somalia	14	637657	7691	Africa	3.00	146
## 58	South Africa	32	1221037	36070	Africa	6.05	3741
## 59	Sri Lanka	7	65610	17240	Asia	9.59	838
## 60	Sudan	134	2505813	25941	Africa	4.02	435
## 61	Suriname	17	163265	429	Americas	12.00	1903
## 62	Thailand	82	513115	56293	Asia	8.04	2845
## 63	Togo	43	56785	3643	Africa	7.91	389
## 64	UAE	9	83600	1629	Asia	0.83	27198
## 65	Uganda	43	235880	19517	Africa	10.14	348

## 66	Vanuatu	111	12189	163	Oceania	12.00	1621
## 67	Venezuela	40	912050	20226	Americas	7.98	3375
## 68	Vietnam	88	331689	68183	Asia	8.80	276
## 69	Yemen	6	527968	12302	Asia	0.00	387
## 70	Zambia	38	752618	8780	Africa	5.43	417
## 71	Zimbabwe	18	390759	10019	Africa	5.29	846

Step 4: Keep rows that are incomplete.

You can also use `complete.cases()` with `filter()` to keep only rows that are incomplete:

```
joinDF %>%
  filter(! complete.cases(joinDF))
```

##	Country	Langs	Area	Population	Continent	GS
## 1	French Guiana	11	90000	102	Americas	10.40
## 2	Tanzania	131	945087	28359	Africa	7.02
## 3	Zaire	219	2344858	36672	Africa	9.44
## 4	Afghanistan	NA	NA	NA	<NA>	NA
## 5	Albania	NA	NA	NA	<NA>	NA
## 6	Andorra	NA	NA	NA	<NA>	NA
## 7	Anguilla	NA	NA	NA	<NA>	NA
## 8	Antigua and Barbuda	NA	NA	NA	<NA>	NA
## 9	Argentina	NA	NA	NA	<NA>	NA
## 10	Armenia	NA	NA	NA	<NA>	NA
## 11	Aruba	NA	NA	NA	<NA>	NA
## 12	Austria	NA	NA	NA	<NA>	NA
## 13	Azerbaijan	NA	NA	NA	<NA>	NA
## 14	Bahamas	NA	NA	NA	<NA>	NA
## 15	Bahrain	NA	NA	NA	<NA>	NA
## 16	Barbados	NA	NA	NA	<NA>	NA
## 17	Belarus	NA	NA	NA	<NA>	NA
## 18	Belgium	NA	NA	NA	<NA>	NA
## 19	Belize	NA	NA	NA	<NA>	NA
## 20	Bermuda	NA	NA	NA	<NA>	NA
## 21	Bhutan	NA	NA	NA	<NA>	NA
## 22	Bosnia and Herzegovina	NA	NA	NA	<NA>	NA
## 23	British Virgin Islands	NA	NA	NA	<NA>	NA
## 24	Brunei Darussalam	NA	NA	NA	<NA>	NA
## 25	Bulgaria	NA	NA	NA	<NA>	NA
## 26	Burundi	NA	NA	NA	<NA>	NA
## 27	Cabo Verde	NA	NA	NA	<NA>	NA
## 28	Canada	NA	NA	NA	<NA>	NA
## 29	Cayman Islands	NA	NA	NA	<NA>	NA
## 30	Chile	NA	NA	NA	<NA>	NA
## 31	China	NA	NA	NA	<NA>	NA
## 32	China, Hong Kong SAR	NA	NA	NA	<NA>	NA
## 33	China, Macao SAR	NA	NA	NA	<NA>	NA
## 34	Comoros	NA	NA	NA	<NA>	NA
## 35	Cook Islands	NA	NA	NA	<NA>	NA
## 36	Croatia	NA	NA	NA	<NA>	NA
## 37	Cyprus	NA	NA	NA	<NA>	NA
## 38	Czechia	NA	NA	NA	<NA>	NA
## 39	Dem. People's Rep. Korea	NA	NA	NA	<NA>	NA
## 40	Dem. Rep. of the Congo	NA	NA	NA	<NA>	NA

## 41	Denmark	NA	NA	NA	<NA>	NA
## 42	Djibouti	NA	NA	NA	<NA>	NA
## 43	Dominica	NA	NA	NA	<NA>	NA
## 44	Dominican Republic	NA	NA	NA	<NA>	NA
## 45	El Salvador	NA	NA	NA	<NA>	NA
## 46	Equatorial Guinea	NA	NA	NA	<NA>	NA
## 47	Eritrea	NA	NA	NA	<NA>	NA
## 48	Estonia	NA	NA	NA	<NA>	NA
## 49	Eswatini	NA	NA	NA	<NA>	NA
## 50	Fiji	NA	NA	NA	<NA>	NA
## 51	Finland	NA	NA	NA	<NA>	NA
## 52	France	NA	NA	NA	<NA>	NA
## 53	French Polynesia	NA	NA	NA	<NA>	NA
## 54	Gambia	NA	NA	NA	<NA>	NA
## 55	Georgia	NA	NA	NA	<NA>	NA
## 56	Germany	NA	NA	NA	<NA>	NA
## 57	Greece	NA	NA	NA	<NA>	NA
## 58	Greenland	NA	NA	NA	<NA>	NA
## 59	Grenada	NA	NA	NA	<NA>	NA
## 60	Guinea-Bissau	NA	NA	NA	<NA>	NA
## 61	Haiti	NA	NA	NA	<NA>	NA
## 62	Hungary	NA	NA	NA	<NA>	NA
## 63	Iceland	NA	NA	NA	<NA>	NA
## 64	Iran (Islamic Republic of)	NA	NA	NA	<NA>	NA
## 65	Iraq	NA	NA	NA	<NA>	NA
## 66	Ireland	NA	NA	NA	<NA>	NA
## 67	Israel	NA	NA	NA	<NA>	NA
## 68	Italy	NA	NA	NA	<NA>	NA
## 69	Jamaica	NA	NA	NA	<NA>	NA
## 70	Japan	NA	NA	NA	<NA>	NA
## 71	Jordan	NA	NA	NA	<NA>	NA
## 72	Kazakhstan	NA	NA	NA	<NA>	NA
## 73	Kiribati	NA	NA	NA	<NA>	NA
## 74	Kosovo	NA	NA	NA	<NA>	NA
## 75	Kuwait	NA	NA	NA	<NA>	NA
## 76	Kyrgyzstan	NA	NA	NA	<NA>	NA
## 77	Latvia	NA	NA	NA	<NA>	NA
## 78	Lebanon	NA	NA	NA	<NA>	NA
## 79	Lesotho	NA	NA	NA	<NA>	NA
## 80	Liechtenstein	NA	NA	NA	<NA>	NA
## 81	Lithuania	NA	NA	NA	<NA>	NA
## 82	Luxembourg	NA	NA	NA	<NA>	NA
## 83	Maldives	NA	NA	NA	<NA>	NA
## 84	Malta	NA	NA	NA	<NA>	NA
## 85	Marshall Islands	NA	NA	NA	<NA>	NA
## 86	Mauritius	NA	NA	NA	<NA>	NA
## 87	Micronesia (Fed. States of)	NA	NA	NA	<NA>	NA
## 88	Monaco	NA	NA	NA	<NA>	NA
## 89	Mongolia	NA	NA	NA	<NA>	NA
## 90	Montenegro	NA	NA	NA	<NA>	NA
## 91	Montserrat	NA	NA	NA	<NA>	NA
## 92	Morocco	NA	NA	NA	<NA>	NA
## 93	Nauru	NA	NA	NA	<NA>	NA
## 94	Netherlands	NA	NA	NA	<NA>	NA

## 95	Netherlands Antilles [former]	NA	NA	NA	<NA>	NA
## 96	New Caledonia	NA	NA	NA	<NA>	NA
## 97	New Zealand	NA	NA	NA	<NA>	NA
## 98	North Macedonia	NA	NA	NA	<NA>	NA
## 99	Norway	NA	NA	NA	<NA>	NA
## 100	Pakistan	NA	NA	NA	<NA>	NA
## 101	Palau	NA	NA	NA	<NA>	NA
## 102	Poland	NA	NA	NA	<NA>	NA
## 103	Portugal	NA	NA	NA	<NA>	NA
## 104	Puerto Rico	NA	NA	NA	<NA>	NA
## 105	Qatar	NA	NA	NA	<NA>	NA
## 106	Republic of Korea	NA	NA	NA	<NA>	NA
## 107	Republic of Moldova	NA	NA	NA	<NA>	NA
## 108	Romania	NA	NA	NA	<NA>	NA
## 109	Russian Federation	NA	NA	NA	<NA>	NA
## 110	Rwanda	NA	NA	NA	<NA>	NA
## 111	Saint Kitts and Nevis	NA	NA	NA	<NA>	NA
## 112	Saint Lucia	NA	NA	NA	<NA>	NA
## 113	Saint Vincent & Grenadines	NA	NA	NA	<NA>	NA
## 114	Samoa	NA	NA	NA	<NA>	NA
## 115	San Marino	NA	NA	NA	<NA>	NA
## 116	Sao Tome and Principe	NA	NA	NA	<NA>	NA
## 117	Serbia	NA	NA	NA	<NA>	NA
## 118	Seychelles	NA	NA	NA	<NA>	NA
## 119	Singapore	NA	NA	NA	<NA>	NA
## 120	Slovakia	NA	NA	NA	<NA>	NA
## 121	Slovenia	NA	NA	NA	<NA>	NA
## 122	Spain	NA	NA	NA	<NA>	NA
## 123	State of Palestine	NA	NA	NA	<NA>	NA
## 124	Sweden	NA	NA	NA	<NA>	NA
## 125	Switzerland	NA	NA	NA	<NA>	NA
## 126	Syrian Arab Republic	NA	NA	NA	<NA>	NA
## 127	Tajikistan	NA	NA	NA	<NA>	NA
## 128	Timor-Leste	NA	NA	NA	<NA>	NA
## 129	Tonga	NA	NA	NA	<NA>	NA
## 130	Trinidad and Tobago	NA	NA	NA	<NA>	NA
## 131	Tunisia	NA	NA	NA	<NA>	NA
## 132	Turkey	NA	NA	NA	<NA>	NA
## 133	Turkmenistan	NA	NA	NA	<NA>	NA
## 134	Turks and Caicos Islands	NA	NA	NA	<NA>	NA
## 135	Tuvalu	NA	NA	NA	<NA>	NA
## 136	Ukraine	NA	NA	NA	<NA>	NA
## 137	United Kingdom	NA	NA	NA	<NA>	NA
## 138	United Rep. of Tanzania	NA	NA	NA	<NA>	NA
## 139	United States of America	NA	NA	NA	<NA>	NA
## 140	Uruguay	NA	NA	NA	<NA>	NA
## 141	Uzbekistan	NA	NA	NA	<NA>	NA
## 142	Zanzibar	NA	NA	NA	<NA>	NA
##	GDP					
## 1	NA					
## 2	NA					
## 3	NA					
## 4	189					
## 5	770					

## 6	23359
## 7	10583
## 8	7841
## 9	7993
## 10	426
## 11	16442
## 12	30169
## 13	396
## 14	18264
## 15	12040
## 16	8585
## 17	1412
## 18	28427
## 19	2834
## 20	40780
## 21	562
## 22	529
## 23	21114
## 24	17650
## 25	1723
## 26	168
## 27	1434
## 28	20594
## 29	40930
## 30	5253
## 31	592
## 32	23559
## 33	18139
## 34	1016
## 35	5588
## 36	4849
## 37	15102
## 38	5771
## 39	222
## 40	215
## 41	35356
## 42	788
## 43	3639
## 44	1995
## 45	1590
## 46	427
## 47	207
## 48	3086
## 49	1979
## 50	2570
## 51	26215
## 52	26707
## 53	20542
## 54	1115
## 55	539
## 56	31898
## 57	12687
## 58	23671
## 59	2775

## 60	703
## 61	345
## 62	4486
## 63	26223
## 64	1888
## 65	172
## 66	18958
## 67	18793
## 68	20449
## 69	2592
## 70	43118
## 71	1488
## 72	1282
## 73	720
## 74	2351
## 75	16487
## 76	327
## 77	2154
## 78	3793
## 79	542
## 80	87414
## 81	1848
## 82	52894
## 83	2229
## 84	9722
## 85	2379
## 86	3786
## 87	2063
## 88	100041
## 89	730
## 90	1958
## 91	6728
## 92	1430
## 93	3496
## 94	29240
## 95	14734
## 96	19024
## 97	17184
## 98	2373
## 99	34813
## 100	629
## 101	6398
## 102	3696
## 103	11639
## 104	11721
## 105	15849
## 106	12277
## 107	407
## 108	1640
## 109	2713
## 110	208
## 111	6987
## 112	4199
## 113	2887

```
## 114 1151
## 115 39407
## 116 780
## 117 2208
## 118 8005
## 119 25265
## 120 3713
## 121 10683
## 122 15366
## 123 1254
## 124 29915
## 125 48817
## 126 944
## 127 211
## 128 503
## 129 2115
## 130 4235
## 131 2172
## 132 3985
## 133 520
## 134 12425
## 135 1274
## 136 990
## 137 23042
## 138 260
## 139 28758
## 140 6609
## 141 589
## 142 235
```

Step 5: Find and replace incomplete data.

If you want to replace missing values with the average of those values from the existing data, you can do so by, first, finding the average:

```
# calculate the average GDP
mg <- mean(joinDF$GDP, na.rm = TRUE)
mg
```

```
## [1] 7916.948
```

And, second, using the `is.na()` function to find NA values and replace them with the mean:

```
joinDF[is.na(joinDF$GDP), "GDP"] <- mg
head(joinDF)
```

```
##      Country Langs      Area Population Continent  GS  GDP
## 1   Algeria     18 2381741      25660    Africa 6.60 1452
## 2   Angola      42 1246700      10303    Africa 6.22  466
## 3 Australia    234 7713364      17336   Oceania 6.00 21635
## 4 Bangladesh   37  143998     118745     Asia 7.40   319
## 5    Benin      52  112622       4889    Africa 7.14   397
## 6   Bolivia    38 1098581       7612  Americas 6.92   887
```