

# Exploration of Data

Trinity Bibler

2022-07-31

## Summary Statistics and Cleaning the Data

```
drug_df <- read_excel("drugoverdoses.xlsx")

drug_df_clean <- drug_df %>%
  rename(deaths_20 = `Deaths - 2020`,
         population_20 = `Population - 2020`,
         crude_rate_20 = `Crude Rate - 2020`,
         age_adjusted_20 = `Age-Adjusted Rate - 2020`,
         deaths_19 = `Deaths - 2019`,
         population_19 = `Population - 2019`,
         crude_rate_19 = `Crude Rate - 2019`,
         age_adjusted_19 = `Age-Adjusted Rate - 2019`,
         deaths_18 = `Deaths - 2018`,
         population_18 = `Population - 2018`,
         crude_rate_18 = `Crude Rate - 2018`,
         age_adjusted_18 = `Age-Adjusted Rate - 2018`,
         deaths_17 = `Deaths - 2017`,
         population_17 = `Population - 2017`,
         crude_rate_17 = `Crude Rate - 2017`,
         age_adjusted_17 = `Age-Adjusted Rate - 2017`,
         deaths_16 = `Deaths - 2016`,
         population_16 = `Population - 2016`,
         crude_rate_16 = `Crude Rate - 2016`,
         age_adjusted_16 = `Age-Adjusted Rate - 2016`
  )

drug_deaths <- drug_df_clean %>%
  select(deaths_20, deaths_19, deaths_18, deaths_17, deaths_16)

drug_deaths_clean <- drug_deaths[-c(52:56),]

transform(drug_deaths_clean, deaths_20 = as.numeric(deaths_20))
```

	deaths_20	deaths_19	deaths_18	deaths_17	deaths_16
## 1	5508	3235	2410	2199	2012
## 2	5470	3771	3189	3245	2798
## 3	4385	3452	3237	4293	3613
## 4	4233	2939	2991	3224	3009
## 5	3907	3034	2866	2548	2235
## 6	2948	2219	2169	2201	1947
## 7	2634	1817	1783	1953	1505

## 8	2538	2502	2583	1969	1409
## 9	2504	2104	2087	1985	1821
## 10	2412	1543	1307	1269	1186
## 11	2190	1789	2011	2033	1762
## 12	2124	1497	1402	1458	1375
## 13	2065	1969	1991	1913	1990
## 14	1897	1266	1193	1241	1130
## 15	1884	1290	1106	927	769
## 16	1877	1254	1104	1176	794
## 17	1688	1036	989	1160	989
## 18	1409	862	835	748	628
## 19	1377	1096	1132	952	914
## 20	1305	862	866	1014	918
## 21	1251	928	846	926	865
## 22	1250	1100	948	955	855
## 23	1195	827	737	742	709
## 24	1129	673	702	833	733
## 25	978	633	564	578	536
## 26	945	558	444	415	346
## 27	678	428	343	422	396
## 28	611	414	381	422	343
## 29	559	373	372	412	408
## 30	535	394	338	332	349
## 31	499	332	339	344	312
## 32	447	397	437	456	466
## 33	420	323	282	360	301
## 34	407	387	355	250	154
## 35	405	245	173	185	180
## 36	341	365	412	424	437
## 37	337	243	191	244	209
## 38	331	240	267	277	279
## 39	320	268	308	388	444
## 40	284	198	208	188	169
## 41	262	184	156	144	146
## 42	224	161	143	206	183
## 43	160	133	120	103	119
## 44	157	114	127	114	101
## 45	112	83	68	102	94
## 46	101	68	63	59	44
## 47	86	70	64	38	42
## 48	74	55	59	53	77
## 49	71	43	36	35	54
## 50	60	47	40	47	50
## 51	46	39	28	35	42

```
summary <- drug_deaths_clean %>%
  summarize_all(list(mean, median, sd, IQR, min, max))
```

```
## Warning in mean.default(deaths_20): argument is not numeric or logical:
## returning NA
```

```
summary_clean <- summary %>%
  rename()
```

```
#Writing Clean Table to CSV file
```

```
drug_deaths_by_state <- drug_df_clean %>%  
  select(State, deaths_20, deaths_19, deaths_18, deaths_17, deaths_16)  
  
drug_deaths_by_state2 <- drug_deaths_by_state[-c(52:56),]  
  
write.csv(drug_deaths_by_state2, file = "drug_deaths.csv")
```