

DATA 202 Homework 4

Adham Rishmawi

2022-10-23

```
##importing data
```

```
data1 <- read.csv("~/data202/homework4/country-population.csv")
data2 <- read.csv("~/data202/homework4/country-codes.csv")
```

```
##merging datasets
```

```
joined_data <- merge(data1, data2, by.x = "country",
                      all.x = TRUE, all.y = FALSE)
```

```
####filtering bad instances
```

```
df2 <- joined_data[,!names(joined_data) %in% c("code")]
```

```
##Query Raw Population Data
```

```
staff4 <-df2 |>
  group_by(`country`, `X1800`, `X1880`)
```

```
##Exploring the data
```

Initial question: do populations in European countries increase at a less rate than Asian countries

plot

We added a column by mutating a new one in and calling it change and making it the difference between the last recorded population and the first recorded

```
staff4 <-df2 |>
  group_by(`country`, `X1800`, `X1880`) |>
  mutate(change = `X2015` - `X1800`) |>
  mutate(chnage_upper = `X2015` - `X2010`) |>
  mutate(change_lower = `X1850` - `X1800`)
```

```
staff4
```

```
# A tibble: 275 x 85
```

```
# Groups:   country, X1800, X1880 [231]
```

	country	X1800	X1810	X1820	X1830	X1840	X1850	X1860	X1870	X1880	X1890
	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	afg	3.28e6	3.28e6	3.32e6	3.45e6	3.63e6	3.81e6	3.97e6	4.17e6	4.42e6	4.71e6
2	ago	1.57e6	1.57e6	1.60e6	1.69e6	1.81e6	1.95e6	2.11e6	2.29e6	2.47e6	2.68e6
3	alb	4.10e5	4.24e5	4.39e5	4.57e5	4.78e5	5.07e5	5.53e5	6.10e5	6.73e5	7.42e5
4	and	2.65e3	2.65e3	2.7 e3	2.84e3	3.03e3	3.23e3	3.44e3	3.65e3	3.88e3	4.13e3
5	are	4.02e4	4.02e4	4.02e4	4.02e4	4.02e4	4.02e4	4.02e4	4.09e4	4.27e4	4.52e4
6	arg	5.34e5	5.34e5	5.71e5	6.87e5	8.74e5	1.11e6	1.42e6	1.86e6	2.49e6	3.40e6

```

7 arm      4.13e5 4.13e5 4.24e5 4.54e5 4.97e5 5.44e5 5.96e5 6.52e5 7.14e5 7.81e5
8 atg      3.7 e4 3.7 e4 3.7 e4 3.7 e4 3.7 e4 3.7 e4 3.65e4 3.55e4 3.52e4 3.63e4
9 aus      3.51e5 3.42e5 3.34e5 3.48e5 4.34e5 7.43e5 1.26e6 1.72e6 2.25e6 3.09e6
10 aut     3.21e6 3.29e6 3.39e6 3.54e6 3.73e6 3.96e6 4.24e6 4.56e6 4.95e6 5.41e6
# ... with 265 more rows, and 74 more variables: X1900 <int>, X1910 <int>,
#   X1920 <int>, X1930 <int>, X1940 <int>, X1950 <int>, X1951 <int>,
#   X1952 <int>, X1953 <int>, X1954 <int>, X1955 <int>, X1956 <int>,
#   X1957 <int>, X1958 <int>, X1959 <int>, X1960 <int>, X1961 <int>,
#   X1962 <int>, X1963 <int>, X1964 <int>, X1965 <int>, X1966 <int>,
#   X1967 <int>, X1968 <int>, X1969 <int>, X1970 <int>, X1971 <int>,
#   X1972 <int>, X1973 <int>, X1974 <int>, X1975 <int>, X1976 <int>, ...

```

We Can see that the population increase for a majority of Asian countries has sky rocketed over years and in both sub sections of change_lower and change_higher. This could be due to the fact that Asian countries typically have a larger family size then compared to the ones found in European countries. The change_higher is much smaller in European countries then it is in Asian countries which indicates to the observer of this plot that Asian countries have maintained a higher population in recent years and almost over every year.

Conclusion

My initial question/statement aligns with the truth and I am able to conclude that Asian countries are increasing faster in population then European countries!