Week 2: Data & Variables

POP88162 Introduction to Quantitative Research Methods

Useful mathematical functions in R

```
sum(<numbers>) # Sum
mean(<numbers>) # Mean
median(<numbers>) # Median
sd(<numbers>) # Standard Deviation
log(<number>) # Logarithm
exp(<number>) # Exponent
sqrt(<number>) # Square root
```

Working with data

Download the dataset on inter-state wars from The Correlates of War Project. This dataset is also available on Blackboard under section "Data". The codebook for this dataset is available here.

Read it into R using `read.csv` function or `read_csv` from `readr` package.

```
wars <- read.csv("Your file path")
```

Check the top and bottom of the data.

```
head(wars)
tail(wars, 20)
```

How many observations are there? And how many variables?

```
str(wars)
dim(wars)
nrow(wars)
ncol(wars)
```

What constitutes an individual observation?

Working with variables

Check how many unique wars are listed in the dataset.

```
length(unique(<dataset>$<variable_name>))
```

First, select the variable which records start years of each war (`StartYear1`).

Now select the variable which records end years of each war (`EndYear2`).

Create a new variable, which shows the duration of each war in years by subtracting start years from end years. You can call it something like `war_duration`.

Keep only unique wars in the dataset.

```r
unique_wars <- wars[unique(wars$WarNum),]
```

Calculate the mean and median of this new variable. What is the mean and what is the median duration of inter-state wars in this dataset in years.