CPSC – 375 – 02

Linh Nguyen

**PROJECT 1 REPORT**

1. **Data preparation:**
2. Read data files:

Store dataset that that contains the number of vaccine doses given everyday in different countries as “vaccine\_doses” data frame.

Store dataset that that contains hospital beds as “hospital\_beds” data frame.

Store demographics dataset as “demographics” data frame.

1. Vaccine\_doses data frame:
2. Discard unnecessary data:

* Remove rows under countries that have data of provinces or states:

Rows which are data of province/states are not NA. Rows having data of countries will be NA in column Province\_State. Then remove rows which Province\_State is not NA.

* Only keep columns that have country name, population, number of shots each day. Remove other columns: UID, iso2, iso3, code3, FIPS, Admin2, Province\_States, Lat, Long\_, Combined\_Key.

1. Tidy up table:

Mess type 1: Column headers are values, not variable names.

Solution: pivot\_longer. There will be a column of date and a column of number of vaccine shots.

Remove rows that number of vaccine shots is NA or 0.

1. Calculate required variable:

* Vaccination rate:

Add a new column to the table named “vaccination\_rate”, its value is calculated by dividing the number of vaccine shots by population.

* Day since vaccination starts:

Add a new column called “day\_since\_start”, its value starts with 1 as the first day of vaccination in that country, increasing until the last day of the dataset.

1. Hospital\_bed data frame:

Keep rows that have hospital beds for the most recent 5 years for each country.

1. Demographics data frame:
2. Discard unnecessary data:

Remove `Country Code`, `Series Names` columns.

1. Tidy up table:

Mess type 2: one observation with multiple rows.

Solution: pivot\_wider. There will be columns of population for different age ranges.

1. Add up the male/female population:

Add new columns that store the sum of male and female population of each age range.

Remove the old columns of male/female population.

1. Merge data:

* Match country names between tables:
* “Korea, Rep.” (demographics) replaced by “South Korea”
* “Iran, Islamic Rep.” (demographics) replaced by “Iran”
* “Iran (Islamic Republic of)” (hospital\_beds) replaced by “Iran”
* “United Kingdom of Great Britain and Northern Ireland” (hospital\_bed) replaced by “Unied Kingdom”
* “Republic of Korea” (hospital\_beds) replaced by “South Korea”
* “Korea, South” (vaccine\_doses) replaced by “South Korea”
* Join tables:

Match column name to join: Rename “Country\_Region” to “Country”, “Country Name” to “Country”

Full join vaccine\_doses table and hospital\_beds table by “Country”, then full join the output table with demographics table by “Country”, saved as “my\_data” data frame.

Table

Description automatically generated with medium confidence

1. **Linear modeling the Covid vaccine rate:**