Final.Rmd outlines the final steps that are used to sort vaccine scheduling and assign to dummy data

**Step 0: Load Required Packages**

1. Here: Allows easy access to files and GitHub organizational structure between computers to ensure that when the programs are run, all necessary files are accessible.
2. Tidyverse: allows use of alternative merging and other packages all bundled together
   1. Contains essential packages including Lubridate: the package primarily used to deal with time calculations
      1. Cheat sheet and documentation of package included in “Documentation folder”
3. Readxl: The reading, saving and importing of excel files into R
4. DT: Alternative way to display data tables in R
5. Knitr: Used for Another way to display tables in R
6. Httr: Pulls data from URL download link and imports into R

**Step 1: Load vaccine event (Dummy data)**

1. Located in data folder as DummyData.Rdata
   1. Loaded with here() package
   2. Created in excel based on a format that may be used in Georgia Department of Public Health Data

**Step 2: driver\_table.R**

1. Function requires inputs from user on:
   1. Vaccine df, patient id column, bday column, vaccine date column
2. Creates and standardized column names based on user inputs to:
   1. Vax\_df\_patient\_id\_data
   2. Vax\_df\_bday\_data
      1. Modified to use end of every month like data from GDPH (vax\_df\_bday\_end\_data)
      2. Used in Dummy data to know birth month to start calculations
   3. Vax\_df\_recent\_vax\_date
      1. User inputs column in which the most recent vax date is stored
      2. Recent\_vax finds the date in that column and uses it in the next calcuations
3. Create interval and duration variables for the patient birthday – latest vax event
   1. Time\_diff\_bday\_recent\_vax\_int – interval var
   2. Time\_diff\_bday\_recent\_vax\_dur – converts interval to duration
   3. Time\_diff\_bday\_recent\_vax\_days – converts duration variable to specific days
4. Merge all variables to create dataset vax\_patient\_level\_data
   1. Includes only unique data (no repeats)
5. Load previously created vaccine schedule created previously using here() package
   1. Vaccine schedule was created based off publicly available CDC vaccine recommendations for [children](https://www.cdc.gov/vaccines/hcp/imz-schedules/child-adolescent-age.html) and [adults.](https://www.cdc.gov/vaccines/hcp/imz-schedules/adult-age.html) Dataset includes information on time, dosage amounts and spacing.
   2. Some vaccines have an alternative schedule that can be followed if patient did not adhere to strict CDC schedule (labeled alternative in the original created document)
6. Use latest valid event (recent\_vax) to create list of vaccines that should have been received based on patient birthday and difference between birthday and latest valid event (time\_diff\_bday\_recent\_vax\_days)
   1. Organize imported vaccine schedule to include only On-Time vaccines and remove the alternative scheduling with a filter
   2. Create dose\_admin\_end\_days column to count how many days vaccine was from most recent vaccine event
7. Create merged\_data with vax\_patient\_level\_data creating all possible combinations and filtering them to show those greater than the most recent vax event
   1. Create utd (Up to date) status variables to show the range of the possible times in which vaccination can occur in order to be considered on-time
   2. utd\_range\_start / utd\_range\_end
8. Save and export merged\_data to file data/driver\_table.Rdata

**Step 3: Join vaccine event data to the driver table to determine vaccines actually received (vaccine\_event\_processing.R)**

1. Function requires inputs from user on:
   1. Vaccine df, patient id column, bday column, cvx column, between\_vax\_bday\_column
   2. This function uses the httr package to load data from the CDC website
2. Loaded cvx code dataset with vaccine group from cdc.gov link that includes entire list of current and past cvx codes
   1. Download to a tmp file and name it cdc\_vax\_cvx
3. Subset data to vax\_cvx\_short to avoid unnecessary information
   1. Includes only 3 columns
4. Rename 3 columns to match vaccine schedule used in step 2
5. Load vaccine schedule dataset (vax\_schedule.Rdata)
6. Create patient id, bday, cvx code, and diff\_vax\_dob\_data columns from user inputs in function
7. Merge created columns from user inputs into dataset -> sorted\_vax\_data
8. Merge sorted\_vax\_data and vax\_cvx\_short datasets -> combined\_vax\_data to properly assign vaccine groups with matching cvx codes from user inputted dataset
9. Load driver\_table.Rdata
10. Subset to include only revelant data columns
    1. Vax\_df\_patient\_id\_data
    2. Vaccine\_group
    3. Expected\_vaccine\_dose <- Renamed from Vaccine\_dose
    4. Utd\_range\_start\_days
    5. Utd\_range\_end\_days
11. Load vax\_group\_relationship.Rdata
    1. Merge with combined\_vax\_data to ensure proper merge based on vaccine group names and cvx codes
12. Preform outer merge using full\_join() joining by:
    1. patient id, vaccine group and if vaccine data recieved is between diff\_vax\_dob\_data, utd\_range\_start\_days and utd\_range\_start\_days.
    2. Do this to determine if vaccines received are on time and within the expected vaccine scheduling period based on birthdays
13. View and return dataset: actual\_vaccines\_received which is combination of the datasets from previous function and new datasets

**Step 4: Create Vaccine Timeline and plot it using data from step 3 (timeline.R)**

1. Pull data from previous step to follow data structure
   1. Should be called vaccine\_event\_processed.Rdata
2. Subset data to include only relevant data and vaccines
   1. Vaccines that were dropped included: COVID-19, RSV, Influenza, Hepatitis A

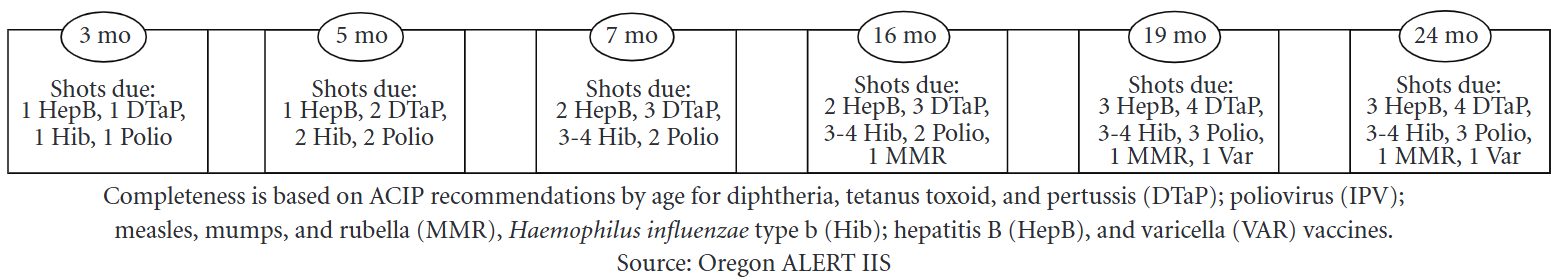
|  |  |
| --- | --- |
| **Variable Name** | **Purpose** |
| vax\_df\_patient\_id\_data | Reports what ID will display in timeline |
| diff\_vax\_dob\_data | Used to calculate estimate DOB |
| Vaccine\_group\_schedule | Name of vaccine according to vaccine schedule excel sheet |
| Vax\_count\_col | Reported vaccine dose |
| Expected\_vaccine\_dose | Dosage based on date in accordance to CDC schedule |
| Utd\_range\_start\_days | Date Range start of up-to-date vaccine |
| Utd\_range\_end\_days | Date range end for up-to-date vaccines |

1. Create actual\_vaccination\_days variable to convert to time in days from previous estimated DOB
2. Create timeline basis with legend using utd\_range\_start\_days and utd\_range\_end\_days for the x-axis and the vaccine\_group\_schedule for the y-axis
   1. The first Geom\_segment adds the black bar that includes the entire expected up-to-date range from start to end
   2. Geom\_point adds a gold dot where the actual point in which a dose was received is set to appear on the timeline
   3. The second and third Geom\_segment adds the red and blue dividers to show when another dosage of the specified vaccine would be given
      1. The red line is for the start period of that specific dose
      2. The blue line is when that specific dose period has passed
   4. Scale\_color\_manual adds the labels to the dividers and ensures the correct colors are added to label them
   5. Labs adds the rest of the labels including the title, x-axis and y-axis, an explanation caption, and places it on the top of the graphic
      1. Also sets the theme to include the classic theming

**New Date Formulation:**

**Robison et Al Method** (2009)(Immunization Milestones: A more comprehensive Picture of Age-Appropriate Vaccination)

* Based off of a milestone (“checkpoint”) system at 3, 5, 7, 16, 19, and 24 months of age
  + All immunization status is checked at each milestone to ensure no child has fallen behind in vaccination
  + If they are, given non-UTD status (Up to date status) and stay until next check
* Up to date status is verified by having received all vaccines in the 4:3:1:3:3:1 series by age 2.
  + 4 DTaP, 3 IPV, 1 MMR, 3 Hib\*, 3 HepB, 1 Varicella
  + \*Depending on when received and product type
* Current standards also include 4+ PCV vaccine (Not included in study)
* AND COVID-19 Vaccine 3+ doses



**Elizabeth Et Al Method** (2005) (Timeliness of Childhood Vaccinations in the United States)

* National Immunization Survey data for US children aged 19 – 35 months
* Measured at 24 months of age, vaccination status for each recommended vaccine

**NIS Survey Data 2023 Method**:

* Uses 4:3:1:3:3:1:4 UTD Status
* 4 DTaP, 3 IPV, 1 MMR, 3 Hib, 3 HepB, 1 Varicella, 4 PCV
* Must receive all at soonest 18 months (latest 36 months?)