DASC 1104 Project Proposal

```
library(tinytex)
```

My blog is accessible at the link: https://nifty-wozniak-e3e1d9.netlify.app

```
dat<- read_csv(here::here("data", "tidytuesday", "data", "2020", "2020-03-24", "tbi_age.csv"
))
```

```
## Parsed with column specification:
## cols(
##   age_group = col_character(),
##   type = col_character(),
##   injury_mechanism = col_character(),
##   number_est = col_double(),
##   rate_est = col_double()
## )
```

```
dat1<- read_csv(here::here("data", "tidytuesday", "data", "2020", "2020-03-24", "tbi_year.csv"
))
```

```
## Parsed with column specification:
## cols(
##   injury_mechanism = col_character(),
##   type = col_character(),
##   year = col_double(),
##   rate_est = col_double(),
##   number_est = col_double()
## )
```

```
dat2<- read_csv(here::here("data", "tidytuesday", "data", "2020", "2020-03-24", "tbi_military.csv"
))
```

```
## Parsed with column specification:
## cols(
##   service = col_character(),
##   component = col_character(),
##   severity = col_character(),
##   diagnosed = col_double(),
##   year = col_double()
## )
```

```
glimpse(dat)
```

```
## Observations: 231
## Variables: 5
## $ age_group        <chr> "0-17", "0-17", "0-17", "0-17", "0-17", "0...
## $ type             <chr> "Emergency Department Visit", "Emergency D...
## $ injury_mechanism <chr> "Motor Vehicle Crashes", "Unintentional Fa...
## $ number_est       <dbl> 47138, 397190, 229236, 55785, NA, 24360, 5...
## $ rate_est         <dbl> 64.1, 539.8, 311.6, 75.8, NA, 33.1, 78.8, ...
```

```
glimpse(dat1)
```

```
## Observations: 216
## Variables: 5
## $ injury_mechanism <chr> "Motor vehicle crashes", "Motor vehicle cr...
## $ type             <chr> "Emergency Department Visit", "Emergency D...
## $ year             <dbl> 2006, 2007, 2008, 2009, 2010, 2011, 2012, ...
## $ rate_est         <dbl> 85.3, 83.8, 83.9, 88.7, 95.3, 98.7, 99.9, ...
## $ number_est       <dbl> 254793, 252459, 254391, 270240, 292942, 30...
```

```
glimpse(dat2)
```

```
## Observations: 450
## Variables: 5
## $ service   <chr> "Army", "Army", "Army", "Army", "Army", "Army", "...
## $ component <chr> "Active", "Active", "Active", "Active", "Active",...
## $ severity  <chr> "Penetrating", "Severe", "Moderate", "Mild", "Not...
## $ diagnosed <dbl> 189, 102, 709, 5896, 122, 33, 26, 177, 1332, 29, ...
## $ year      <dbl> 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2...
```

1. Traumatic Brain Injuries For this project, I am examining the traumatic brain injury (TBI) within the data set tbi_age.csv file from the Tidy Tuesday website. The data set has 216 observations of 5 variables. The variable age_group is a factor of 231 levels that represents the age groups of victims of TBI. The variable type is a factor of 231 levels that represents the where the victim went to treat their injury, or the effect of the injury. The variable injury_mechanism is a factor of 231 levels that represents how the victim obtained their injury. The variable number_set is a factor of 231 levels that represents the estimated observed cases in 2014.The variable rate_est is a factor of 231 levels that represents the where the victim went to treat their injury.

With the second, tbi_year.csv data set from Tidy Tuesday website further examines these injuries. The data set has 216 observations of 5 variables. The variable injury_mechanism is a factor of 216 levels that represents how the injury was received. The variable type is a factor of 216 levels that represents went to treat their injury, or the effect of the injury. The variable year is a factor of 216 levels that represents the year of the injury. The variable rate_est is a factor of 216 levels that rate/100,000 in 2014. The variable number_est is a factor of 216 levels that represents estimated observed cases in each year.

The last data set, tbi_military.csv from Tidy Tuesday specifically delves into brain injuries specifically from military personnel. The data set has 450 observations and 5 variables. The variable service is a factor of 450 levels that represents the military branch of the victim's branch. The variable component is a factor of 450 levels that represents whether the victim is active, guard, or reserve). The variable type is a severity of 450 levels that represents the severity of the injury. The variable diagnosed is a factor of 450 levels that represents the number of diagnoses per year. The variable year is a factor of 450 levels that represents year the injury was observed.

Question 1: What age group is most likely to receive a traumatic brain injury from which injury mechanism? To determine this relationship, I will examine a graph relating the age groups to the different injury mechanisms. As well, I want to observe this in a box plot, allowing easy observation over the median, minimum, and maximum ages these injury mechanisms occur.

Question 2: As the prevalence of TBI has increased in the military branches, has there been any decrease in the amount of concussions diagnosed throughout the years? If not, has severity at least lessened throughout the years? To determine, I will generate a graph to visualize the number of diagnosed and year of diagnoses, as well as a visualization of the severity of the TBI compared to the year of diagnoses.

Question 3: Throughout the different data sets, are deaths from TBI age based or severity based? To determine, I will create 2 graphics for each data set, one comparing the ages of the victims to the deaths in that data set, as well as a graphic between the type or severity of the TBI to the deaths.

```
dat3<- read_csv(here::here("data", "tidytuesday", "data", "2020", "2020-08-18", "plants.csv"
))
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   binomial_name = col_character(),
##   country = col_character(),
##   continent = col_character(),
##   group = col_character(),
##   year_last_seen = col_character(),
##   red_list_category = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
dat4<- read_csv(here::here("data", "tidytuesday", "data", "2020", "2020-08-18", "threats.csv"
))
```

```
## Parsed with column specification:
## cols(
##   binomial_name = col_character(),
##   country = col_character(),
##   continent = col_character(),
##   group = col_character(),
##   year_last_seen = col_character(),
##   red_list_category = col_character(),
##   threat_type = col_character(),
##   threatened = col_double()
## )
```

```
dat5<- read_csv(here::here("data", "tidytuesday", "data", "2020", "2020-08-18", "actions.csv"
))
```

```
## Parsed with column specification:
## cols(
##   binomial_name = col_character(),
##   country = col_character(),
##   continent = col_character(),
```

```
##   group = col_character(),
##   year_last_seen = col_character(),
##   red_list_category = col_character(),
##   action_type = col_character(),
##   action_taken = col_double()
## )
```

```
glimpse(dat3)
```

```
## Observations: 500
## Variables: 24
## $ binomial_name     <chr> "Abutilon pitcairnense", "Acaena exigua",...
## $ country           <chr> "Pitcairn", "United States", "Congo", "Sa...
## $ continent         <chr> "Oceania", "North America", "Africa", "Af...
## $ group             <chr> "Flowering Plant", "Flowering Plant", "Fl...
## $ year_last_seen    <chr> "2000-2020", "1980-1999", "1940-1959", "B...
## $ threat_AA         <dbl> 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1,...
## $ threat_BRU        <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0,...
## $ threat_RCD        <dbl> 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1,...
## $ threat_ISGD       <dbl> 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0,...
## $ threat_EPM        <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ threat_CC         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ threat_HID        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ threat_P          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ threat_TS         <dbl> 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0,...
## $ threat_NSM        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ threat_GE         <dbl> 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,...
## $ threat_NA         <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,...
## $ action_LWP        <dbl> 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0,...
## $ action_SM         <dbl> 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ action_LP         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ action_RM         <dbl> 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,...
## $ action_EA         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ action_NA         <dbl> 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1,...
## $ red_list_category <chr> "Extinct in the Wild", "Extinct", "Extinc...
```

```
glimpse(dat4)
```

```
## Observations: 6,000
## Variables: 8
## $ binomial_name     <chr> "Abutilon pitcairnense", "Abutilon pitcai...
## $ country           <chr> "Pitcairn", "Pitcairn", "Pitcairn", "Pitc...
## $ continent         <chr> "Oceania", "Oceania", "Oceania", "Oceania...
## $ group             <chr> "Flowering Plant", "Flowering Plant", "Fl...
## $ year_last_seen    <chr> "2000-2020", "2000-2020", "2000-2020", "2...
## $ red_list_category <chr> "Extinct in the Wild", "Extinct in the Wi...
## $ threat_type       <chr> "Agriculture & Aquaculture", "Biological ...
## $ threatened        <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,...
```

```
glimpse(dat5)
```

```
## Observations: 3,000
```

```
## Variables: 8
## $ binomial_name     <chr> "Abutilon pitcairnense", "Abutilon pitcai...
## $ country           <chr> "Pitcairn", "Pitcairn", "Pitcairn", "Pitc...
## $ continent         <chr> "Oceania", "Oceania", "Oceania", "Oceania...
## $ group             <chr> "Flowering Plant", "Flowering Plant", "Fl...
## $ year_last_seen    <chr> "2000-2020", "2000-2020", "2000-2020", "2...
## $ red_list_category <chr> "Extinct in the Wild", "Extinct in the Wi...
## $ action_type       <chr> "Land & Water Protection", "Species Manag...
## $ action_taken      <dbl> 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,...
```

2. Plants in Danger For this project, I am examining plants in danger within the data set plants.csv file from the Tidy Tuesday website. The data set has 500 observations of 24 variables. The variable bionomial_name is a factor of 500 levels that represents the species name (Genus + species). The variable country is a factor of 500 levels that represents the country the plant is origin to. The variable continent is a factor of 500 levels that represents the continent the plant originated. The variable group is a factor of 500 levels that represents the taxonomic group of the plant.The variable year_last_seen is a factor of 500 levels that represents the last year the plant was last seen. The next 12 variables all have 500 levels and describe the different threats of the plants (AA - agriculture & aquaculture, BRU- biological resource use, RDC-commercial development, ISGD -invasive species, EPM - energy production and mining, CC- climate change, HID -human intrusions, P- pollution, TS- transportation corridor, NSM- natural systems modifications, GE- geological events, NA- unknown).

The next data set, threats.csv, has 6,000 observations and 5 variables. The variable bionomial_name is factor of 6000 levels that represent the species name (Genus + species). The variable country is factor of 6000 levels that represent the species original location.The variable continent is factor of 6000 levels that represent the species originating continent. The variable group is factor of 6000 levels that represent the taxonomic group the plant belongs to. The variable year_last_seen is factor of 6000 levels the last calendar year the plant was seen.The variable red_list_category is factor of 6000 levels that represent the IUCN Red List category. The variable threat_type is factor of 6000 levels that represent the type of threat the plants are in, from the last data set. The variable threatened is factor of 6000 levels that represent a binary of 0 (action not taken) or 1 (aciton taken).

The last data set, actions.csv has 3000 observations and 8 variables. The variable bionomial_name is factor of 3000 levels that represent the species name (Genus + species). The variable country is factor of 3000 levels that represent the species original location.The variable continent is factor of 3000 levels that represent the species originating continent. The variable group is factor of 3000 levels that represent the taxonomic group the plant belongs to. The variable year_last_seen is factor of 3000 levels the last calendar year the plant was seen.The variable red_list_category is factor of 3000 levels that represent the IUCN Red List category. The variable action_type is factor of 3000 levels that represent the type of action the plant has had.The variable action_taken is factor of 3000 levels that represent a binary of 0 (action not taken) or 1 (action taken).

Question 1: Do countries with significant amounts of plants in danger occupy a lot of the action to maintain the longevity of the plant health? Determine the relationship between country and the different plants as well as the relationship between country and the action type, with graphics describing such relationships.

Question 2: Is there trend in the threat type that came help determine the action type that needs to be enforced? To determine, filter the data to match what the threat was and whether or not action was taken to deter that threat.

Question 3: Was there an event in recent history that wiped out, or put multiple plant species in danger? To determine, I will create a graphic graphing the year the plants were last seen as well as the threat the plants are/were in.