

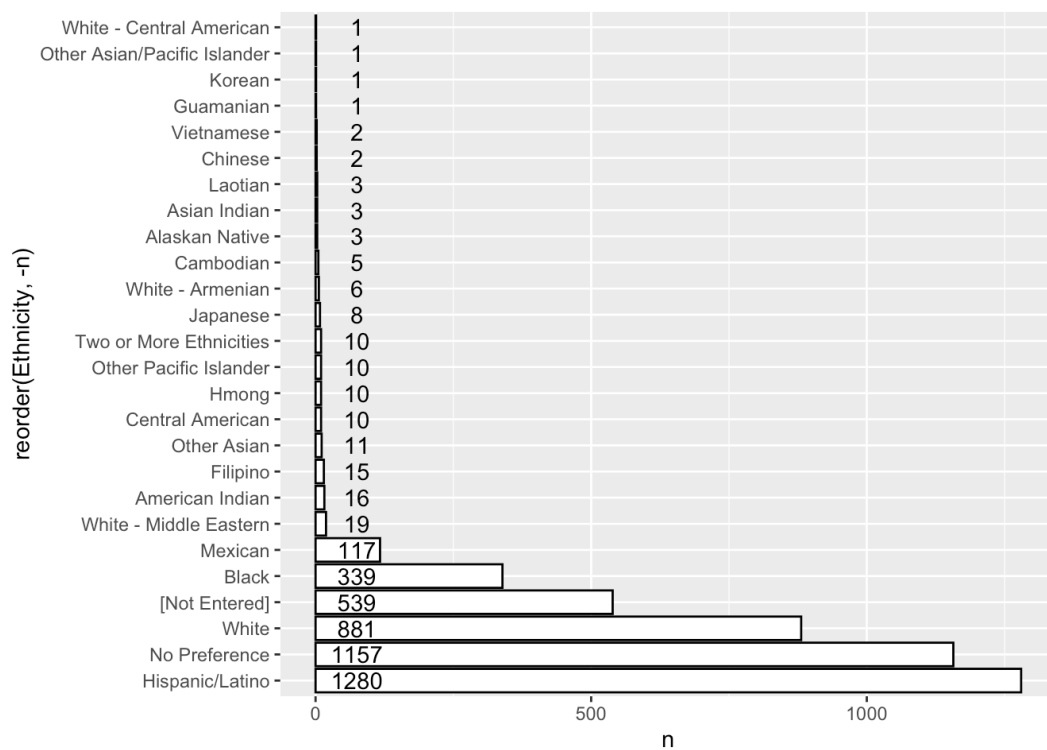
Preliminary Data Analysis

Read in Data

```
— Attaching core tidyverse packages —————
tidyverse 2.0.0 —
✓ dplyr      1.1.4      ✓ readr      2.1.5
✓ forcats    1.0.0      ✓ stringr    1.5.1
✓ ggplot2    3.5.1      ✓ tibble     3.2.1
✓ lubridate  1.9.3      ✓ tidyr      1.3.1
✓ purrr      1.0.2
— Conflicts —————
tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>)
to force all conflicts to become errors
```

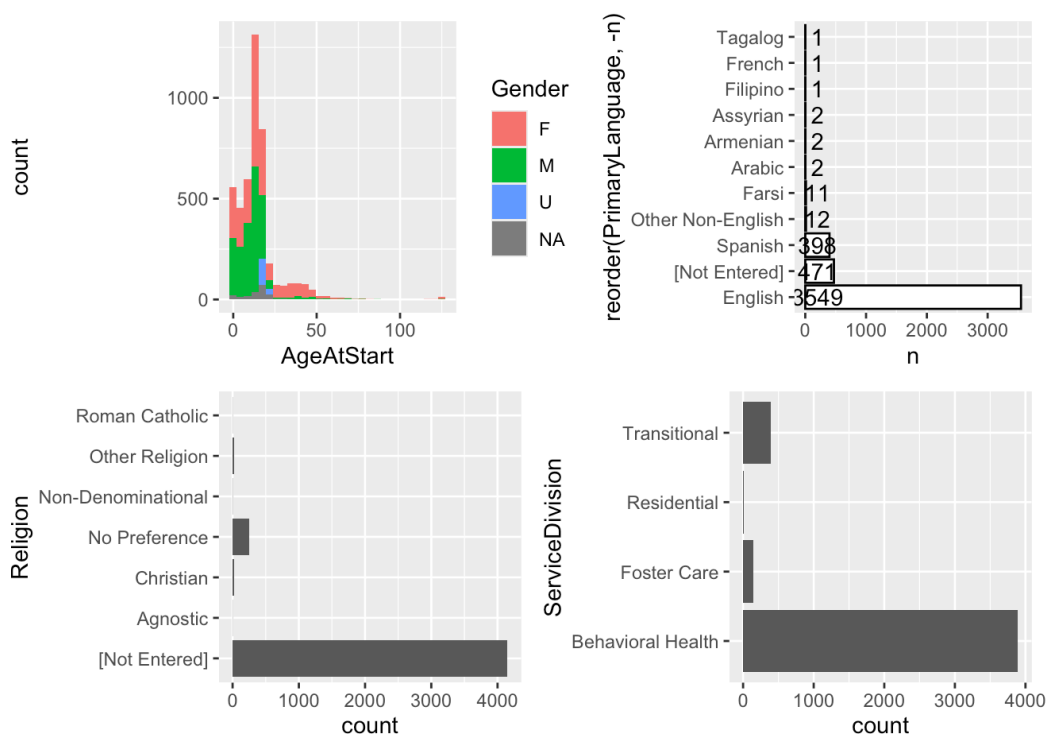
Demographic Analysis

```
demographic_query |>
  count(Ethnicity) |>
  ggplot(aes(reorder(Ethnicity, -n),n)) +
  geom_bar(stat="identity", fill = "white", color = "black") +
  geom_text(aes(label=n),position = position_fill(vjust = 1))
```



```
age <- demographic_query |>
  ggplot(aes(x = AgeAtStart, fill = Gender)) + geom_histogram()
lang <- demographic_query |>
  count(PrimaryLanguage) |>
  ggplot(aes(reorder(PrimaryLanguage, -n), n)) +
  geom_bar(stat="identity", fill = "white", color = "black") +
  geom_text(aes(label=n), position = position_fill(vjust = 1))
religion <- demographic_query |>
  ggplot(aes(x = Religion)) + geom_bar() + coord_flip()
service <- demographic_query |>
  ggplot(aes(x = ServiceDivision)) + geom_bar() + coord_flip()
(age+lang)/(religion+service)
```

`stat_bin()` using `bins = 30`. Pick better value with
`binwidth`.



```
demographic_query |>
  group_by(OptionsNumber) |>
  summarize(n = n()) |>
  arrange(desc(n))
```

A tibble: 3,888 × 2

	OptionsNumber	n
	<dbl>	<int>
1	20048	4
2	81191	4
3	90620	4
4	91843	4
5	92229	4
6	92290	4
7	50147	3
8	54078	3
9	58872	3
10	61853	3

i 3,878 more rows

```
demographic_query |>
  select(OptionsNumber:Religion) |>
  unique() |>
  group_by(OptionsNumber) |>
  mutate(n = n()) |>
  ungroup() |>
```

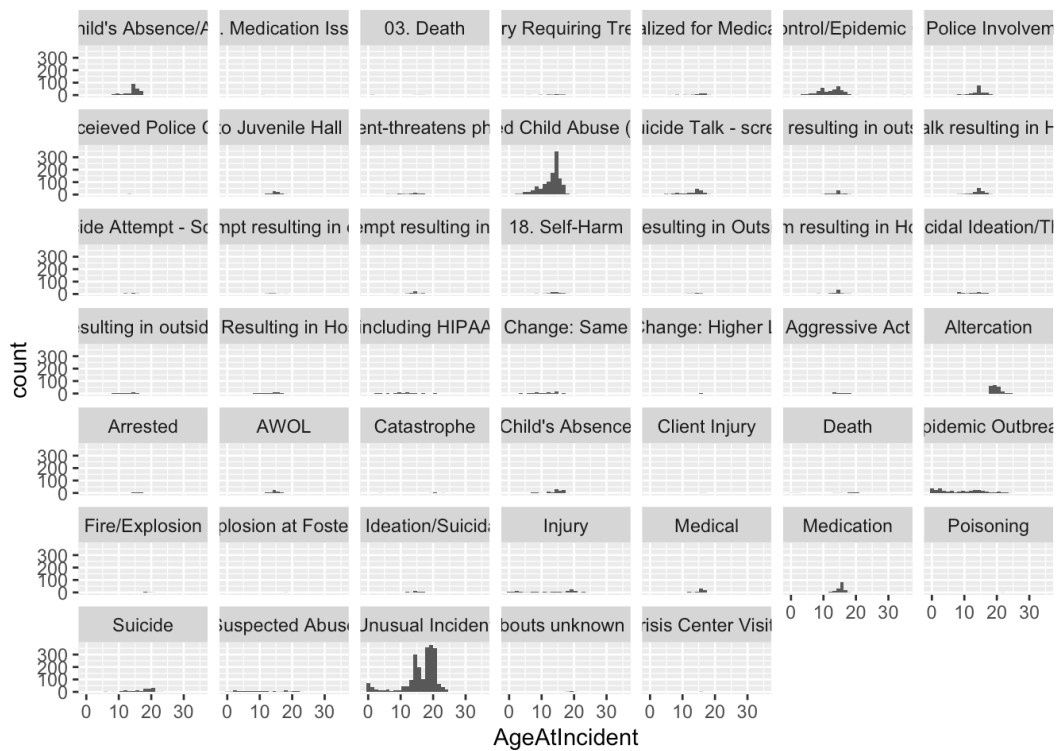
```
arrange(desc(n))
```

```
# A tibble: 4,037 × 7
  OptionsNumber Gender AgeAtStart Ethnicity
PrimaryLanguage Religion      n
      <dbl> <chr>      <dbl> <chr>      <chr>
<chr>      <int>
1         20048 M          15 White      English
[Not Ent...    3
2         20048 M          17 White      English
[Not Ent...    3
3         20048 M          17 White      English
No Prefe...    3
4         61853 M          13 No Preference English
[Not Ent...    3
5         61853 M          15 No Preference English
[Not Ent...    3
6         61853 M          16 No Preference English
[Not Ent...    3
7         86023 M           9 Mexican      English
[Not Ent...    3
8         86023 M          13 Mexican      English
[Not Ent...    3
9         86023 M          15 Mexican      English
[Not Ent...    3
10        20007 F           6 Other Asian  English
[Not Ent...    2
# i 4,027 more rows
```

Incident Analysis

```
incidents |>
  ggplot(aes(x = AgeAtIncident)) + geom_histogram() + f
```

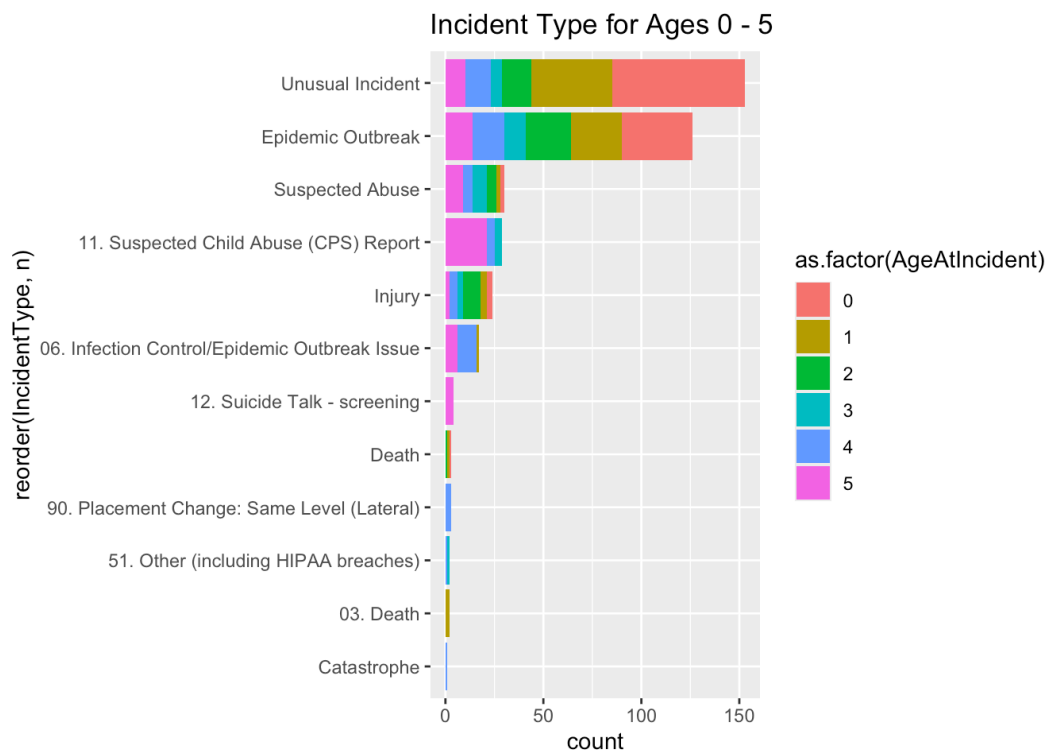
`stat_bin()` using `bins = 30`. Pick better value with
`binwidth`.



```
incidents <- incidents |>
  mutate(AgeRange = case_when(
    AgeAtIncident <= 5 ~ "0-5",
    AgeAtIncident <= 10 ~ "6-9",
    AgeAtIncident <= 15 ~ "10-14",
    AgeAtIncident <= 20 ~ "15-19",
    AgeAtIncident <= 25 ~ "20-24",
    AgeAtIncident <= 30 ~ "25-30",
    AgeAtIncident <= 35 ~ "31-35",
    .default = "35+"))
```

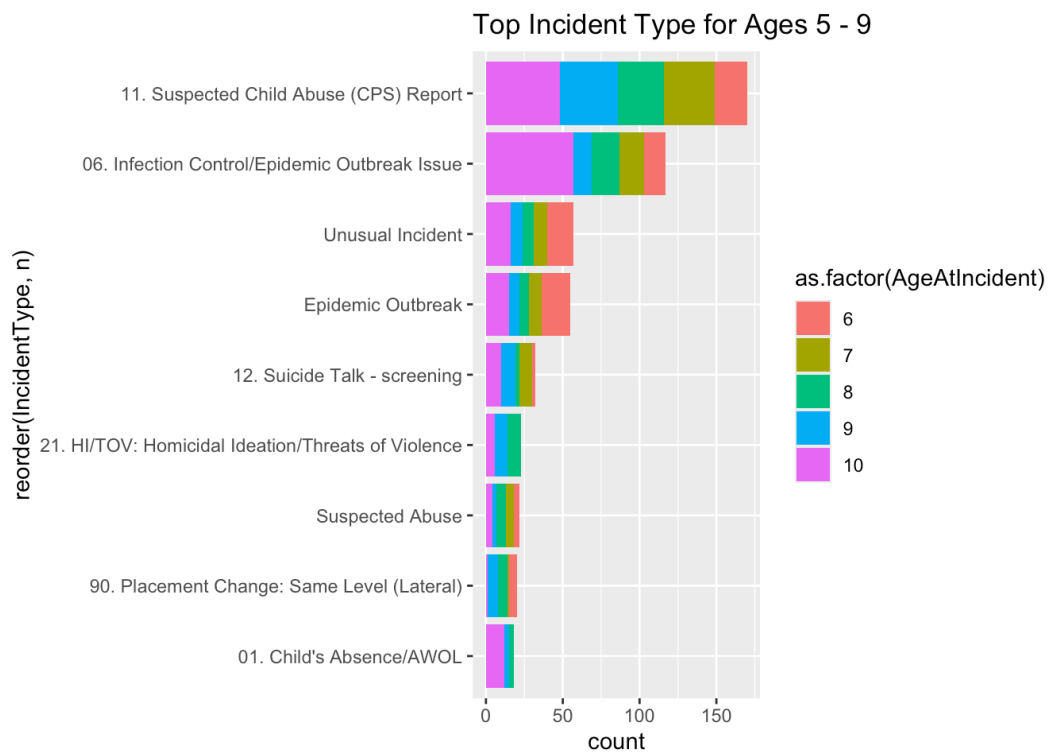
Ages 0 to 5

```
incidents |>
  filter(AgeRange == "0-5") |>
  group_by(IncidentType) |>
  mutate(n = n()) |>
  ungroup() |>
  ggplot(aes(x=reorder(IncidentType,n), fill=as.factor(AgeRange))) +
  geom_bar() + coord_flip() + ggtitle("Incident Type for Ages 0 to 5")
```



Ages 6 to 9

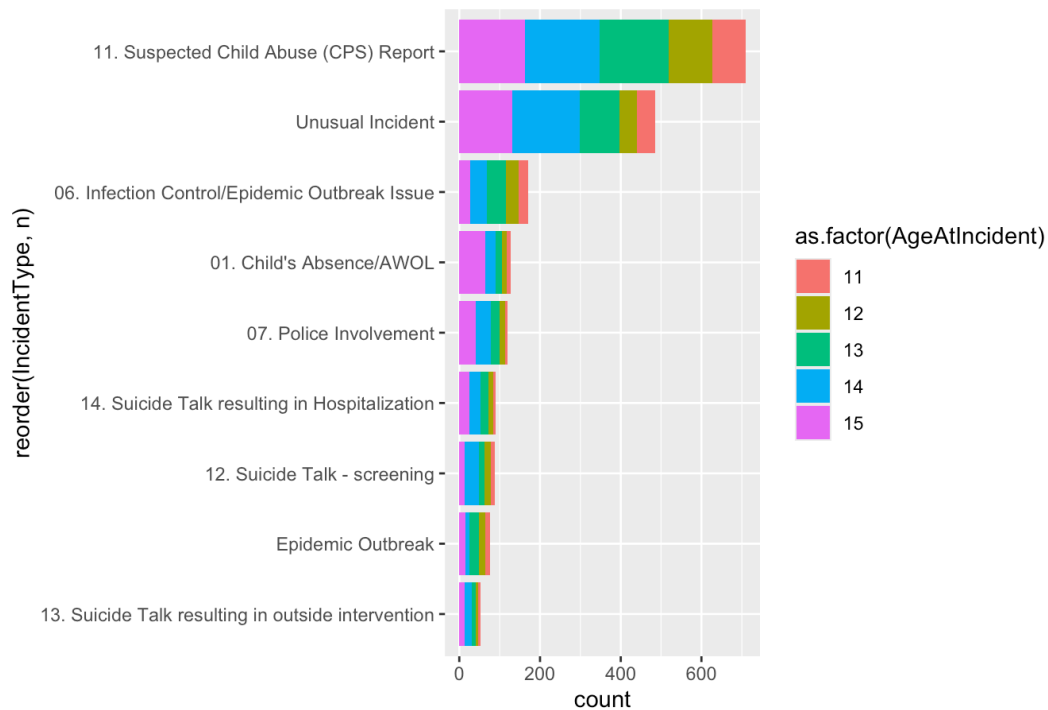
```
incidents |>
  filter(AgeRange == "6-9") |>
  group_by(IncidentType) |>
  mutate(n = n()) |>
  ungroup() |>
  filter(n > 15) |>
  ggplot(aes(x=reorder(IncidentType,n), fill =as.factor
geom_bar() + coord_flip() + ggtitle("Top Incident Typ
```



Ages 10 to 14

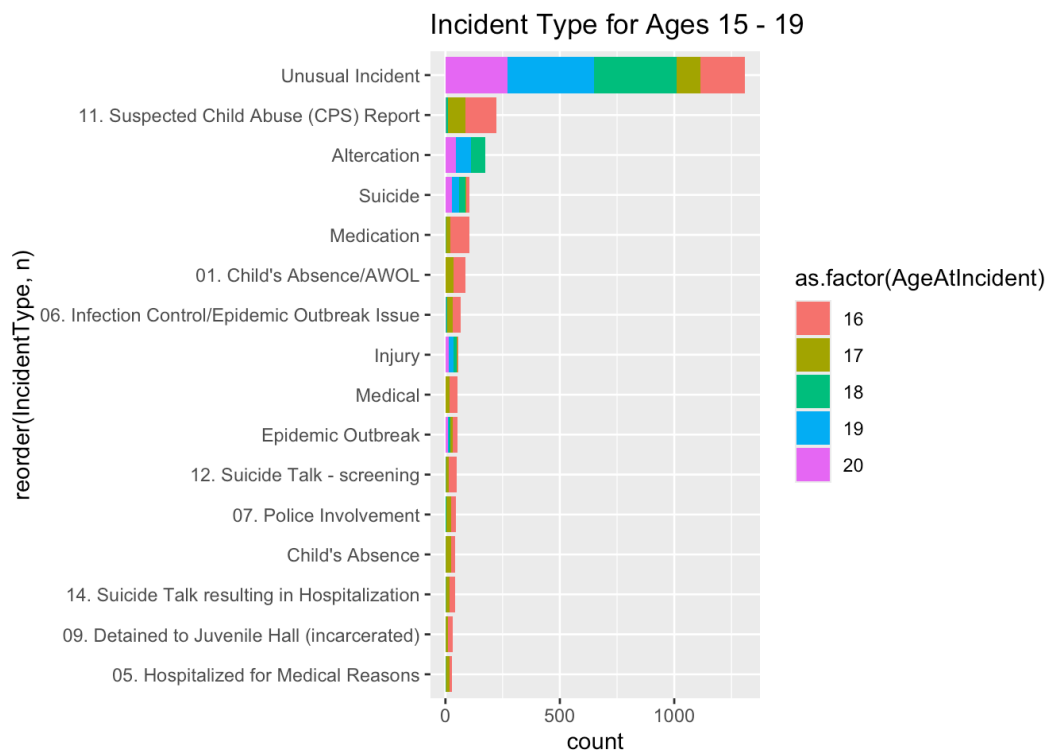
```
incidents |>
  filter(AgeRange == "10-14") |>
  group_by(IncidentType) |>
  mutate(n = n()) |>
  ungroup() |>
  filter(n > 50) |>
  ggplot(aes(x=reorder(IncidentType,n), fill =as.factor(AgeAtIncident))) +
  geom_bar() + coord_flip() + ggtitle("Top Incident Type for Ages 10 to 14")
```

Top Incident Type for Ages 10 - 14



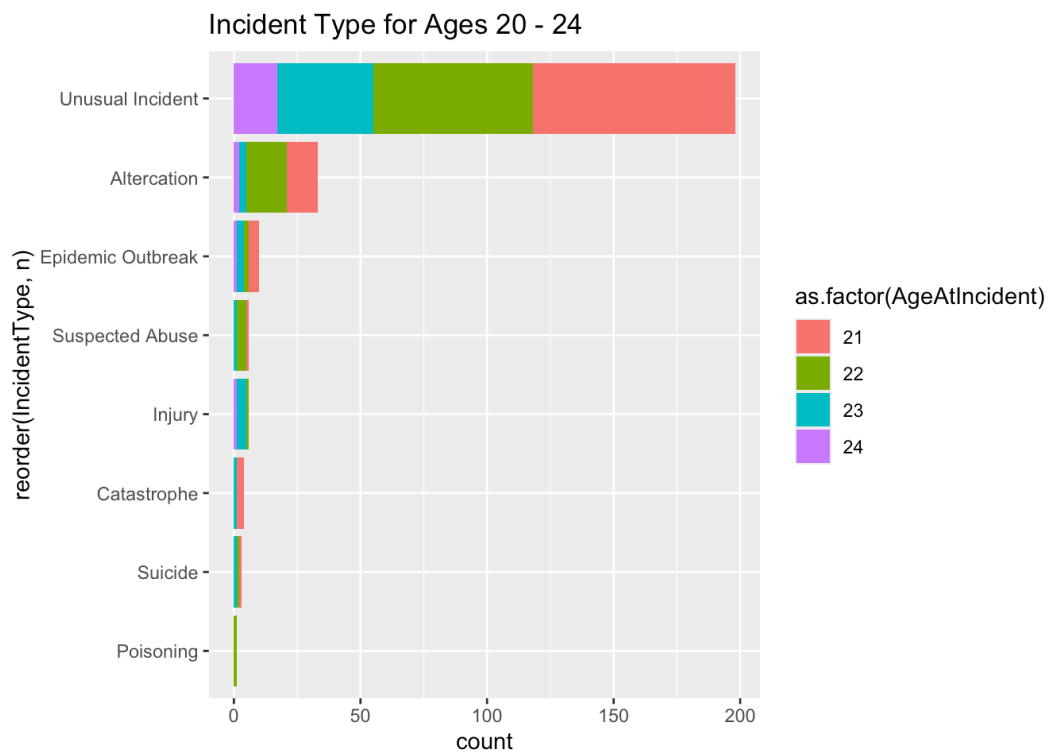
Ages 15 to 19

```
incidents |>
  filter(AgeRange == "15-19") |>
  group_by(IncidentType) |>
  mutate(n = n()) |>
  ungroup() |>
  filter(n > 25) |>
  ggplot(aes(x=reorder(IncidentType,n), fill =as.factor(AgeAtIncident))) +
  geom_bar() + coord_flip() + ggtitle("Incident Type for Ages 15 to 19")
```

Ages 20 to 24

```
incidents |>
  filter(AgeRange == "20-24") |>
  group_by(IncidentType) |>
  mutate(n = n()) |>
  ungroup() |>
  ggplot(aes(x=reorder(IncidentType,n), fill =as.factor
geom_bar() + coord_flip() + ggtitle("Incident Type fo
```



Excluding not major incident types and two clients over the age of 30

```
incidents |>
  group_by(OptionsNumber) |>
  summarize(NumberOfIncidents = n()) |>
  arrange(desc(NumberOfIncidents))
```

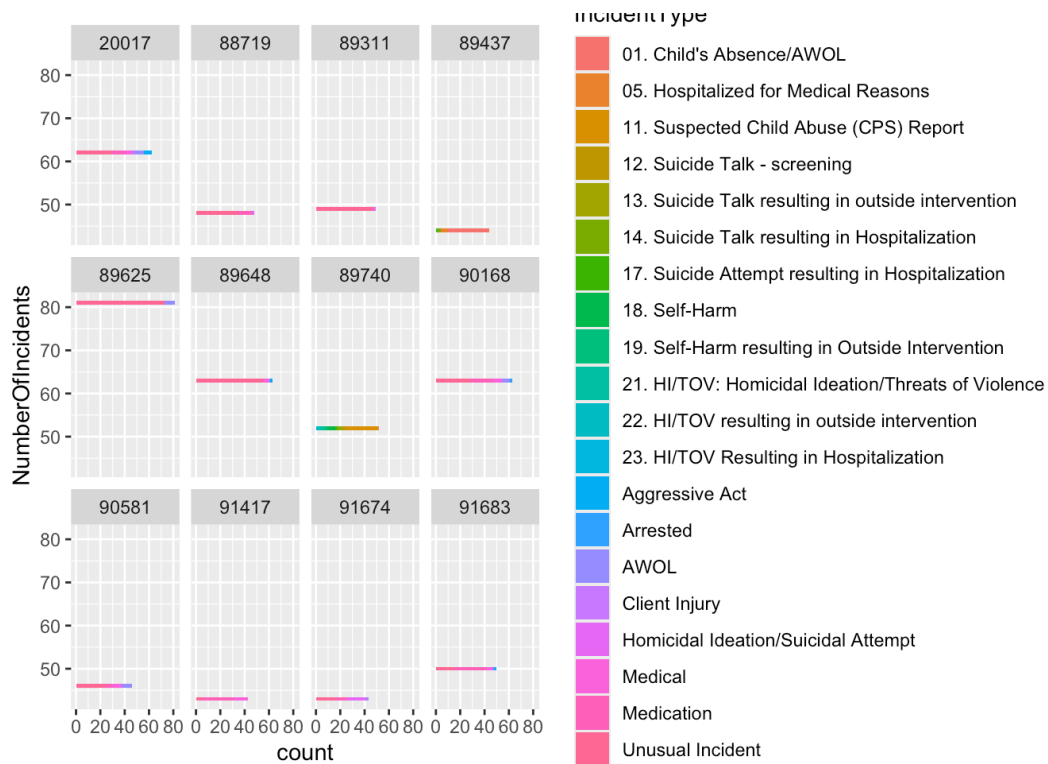
A tibble: 2,117 × 2

	OptionsNumber	NumberOfIncidents
	<dbl>	<int>
1	89625	81
2	89648	63
3	90168	63
4	20017	62
5	89740	52
6	91683	50
7	89311	49
8	88719	48
9	90581	46
10	89437	44

i 2,107 more rows

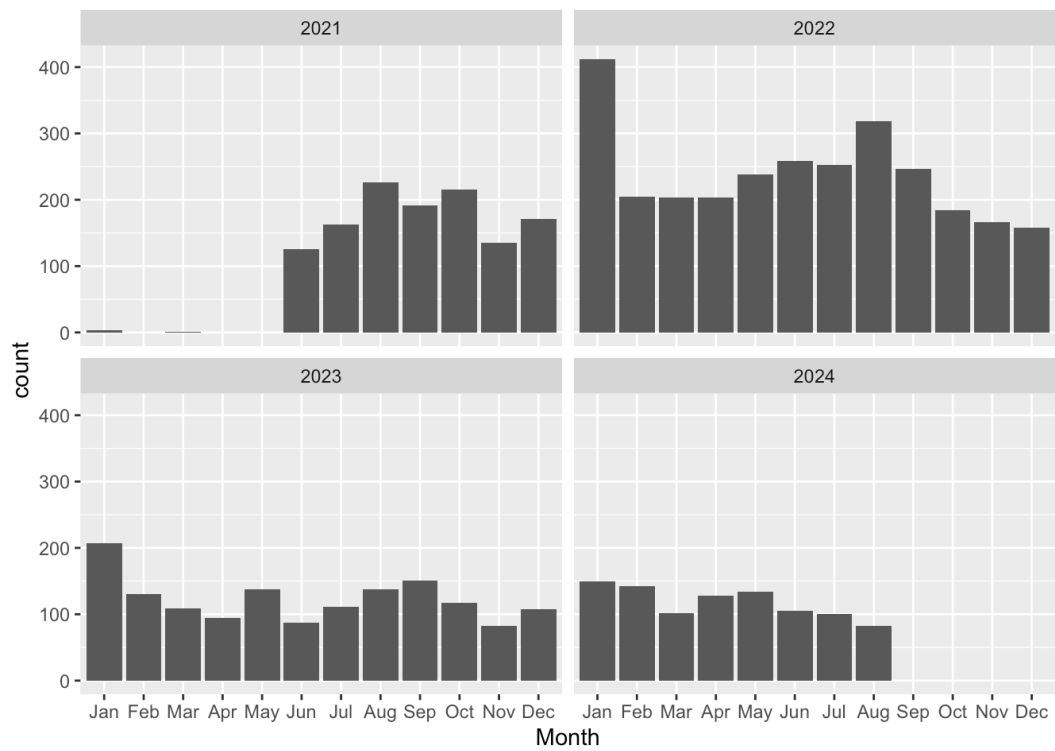
```
incidents |>
  group_by(OptionsNumber) |>
  mutate(NumberOfIncidents = n()) |>
  ungroup() |>
  filter(NumberOfIncidents > 40) |>
```

```
ggplot(aes(x = NumberOfIncidents, fill = IncidentType
```

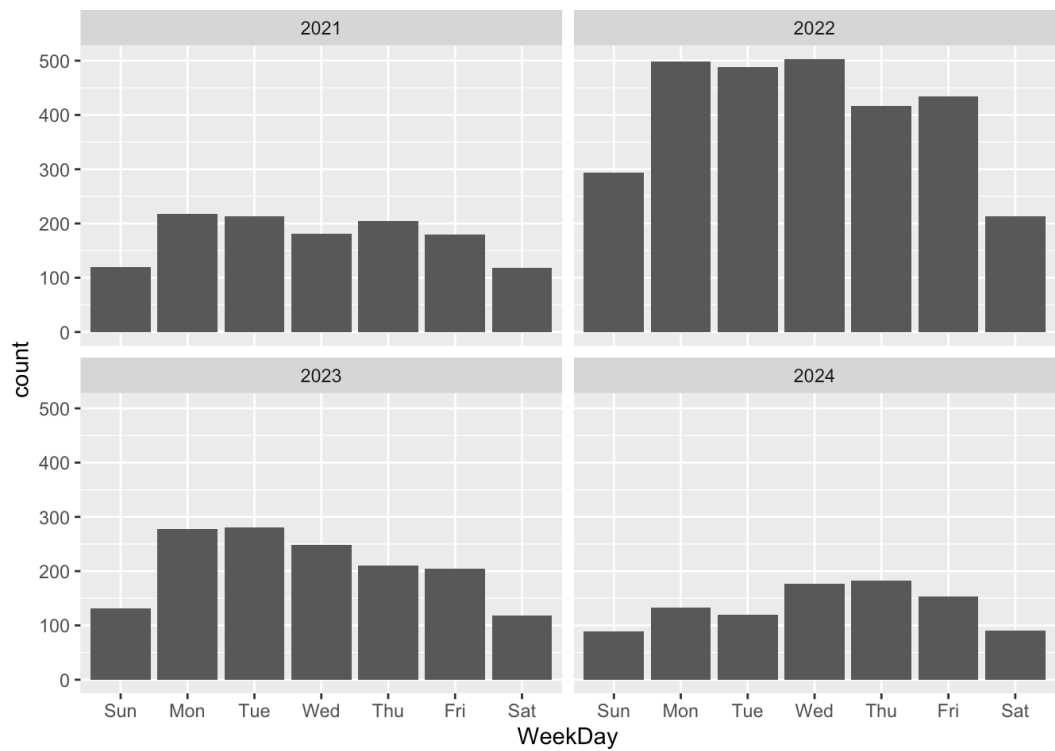


```
incidents <- incidents |>
  mutate(IncidentDate = lubridate::as_datetime(IncidentDate),
         WeekDay = lubridate::wday(IncidentDate, label = TRUE),
         Month = lubridate::month(IncidentDate, label = TRUE),
         Year = lubridate::year(IncidentDate),
         )
```

```
incidents |>
  ggplot(aes(x = Month)) + geom_bar() + facet_wrap(~Year)
```



```
incidents |>
  ggplot(aes(x = WeekDay)) + geom_bar() + facet_wrap(~Y
```



CANS Analysis

```
#cans_last_years
```

Cancelled Appointments Analysis

```
cancelled_apt <- cancelled_apt |>
  mutate(AgeRange = case_when(
    AgeAtAppointment <= 5 ~ "0-5",
    AgeAtAppointment <= 10 ~ "6-9",
    AgeAtAppointment <= 15 ~ "10-14",
    AgeAtAppointment <= 20 ~ "15-19",
    AgeAtAppointment <= 25 ~ "20-24",
    AgeAtAppointment <= 30 ~ "25-30",
    AgeAtAppointment <= 35 ~ "31-35",
    .default = "35+")
cancelled_apt |>
  filter(AgeAtAppointment <= 20) |>
  ggplot(aes(x = AgeRange, fill = AppointmentStatus)) +
  facet_wrap(~Gender) + ggtitle("Appointment Status for
```

