Visualising the RiX Database

This file documents summary statistics of the RiX database, and can be used to indicate data gaps in RiX.

First, load the required libraries and source code:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.3.3     v purrr   0.3.3
## v tibble  3.0.6     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'

## The following object is masked from 'package:purrr':
##
##     set_names

## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(ggplot2)
library(summarytools)
```

```
##
## Attaching package: 'summarytools'

## The following object is masked from 'package:tibble':
##
##     view
```

```
source("./GeneralFunctions.R")
```

```
## Loading required package: sp

##
## Attaching package: 'raster'
```

```
## The following object is masked from 'package:summarytools':
##
##     freq

## The following object is masked from 'package:magrittr':
##
##     extract

## The following object is masked from 'package:dplyr':
##
##     select

## The following object is masked from 'package:tidyr':
##
##     extract

## ### Welcome to rworldmap ###

## For a short introduction type :   vignette('rworldmap')
```

Then load and wrangle the Survey123 database:

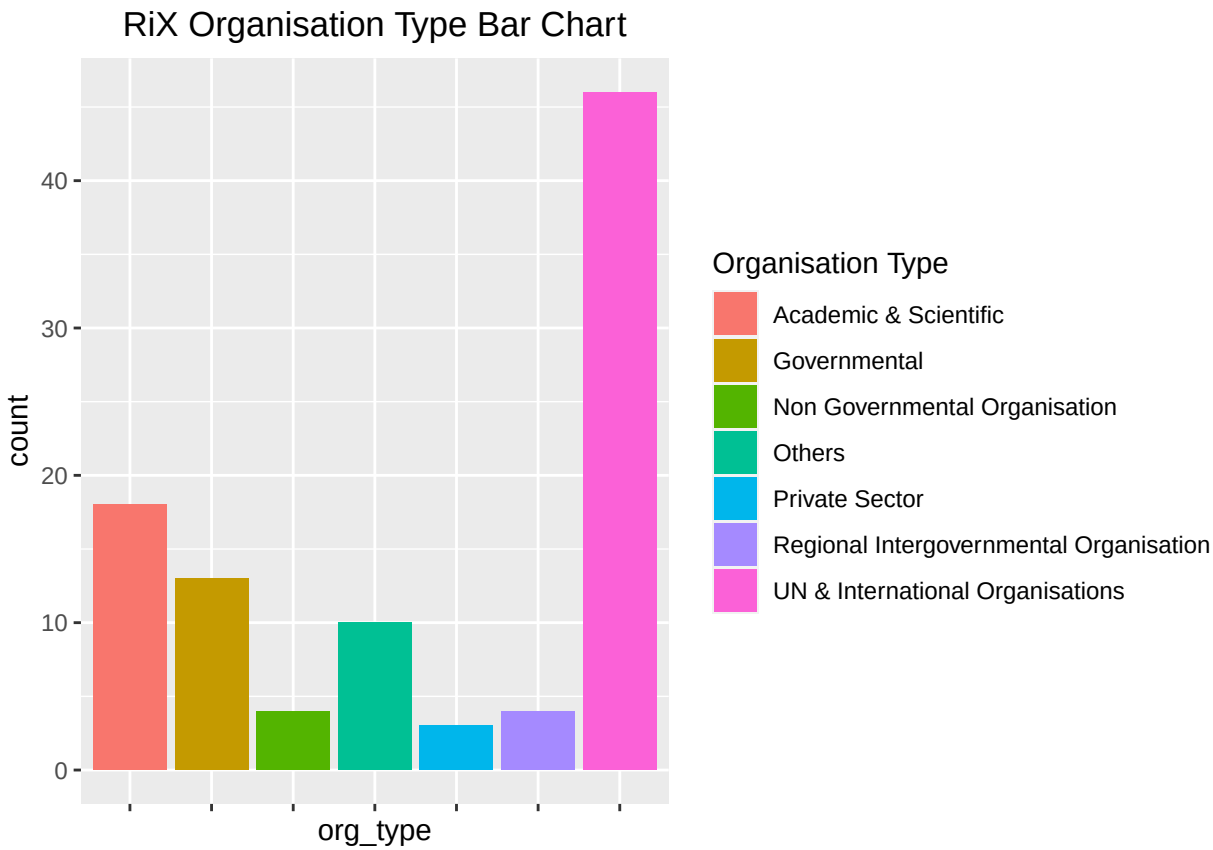
```
rix<-xlsx::read.xlsx("~/Downloads/S123_c1aacc6dde7b4b4b8d43607c1417f283_EXCEL(4).xlsx",
                     sheetName = "Official_RiX_DIIF_XData_Inv_0",as.data.frame = T)

rix%<>%dplyr::select(ObjectID,
                     user,
                     In.English..please.provide.the.name.of.the.dataset,
                     des_org_type,
                     Please.provide.the.name.of.the.organisation.that.produced.the.dataset,
                     Please.provide.the.website.URL.of.the.dataset,
                     des_formattype_dataset,
                     des_formatextension_dataset,
                     Please.provide.the.year.when.the.data.was.published..please.also.provide.the.month
                     des_cca_themes,
                     des_cca_group,
                     des_continent_spatial,
                     concat_countries_spatial,
                     des_final_countries_spatial,
                     final_label_countries,
                     des_resolution_spatial,
                     des_res_spat,
                     What.is.the.starting.year.of.the.data.,
                     What.is.the.end.year.of.the.data..Please.also.provide.the.month.if.known.,
                     Which.of.the.following.hazard.type.s..are.present.in.the.data.,
                     Please.select.hazard.cluster.s..present.in.the.data,
                     Please.select.specific.hazard.s..present.in.the.data,
                     des_exposuretype,
                     des_exposuresubtype,
                     des_vultype,
                     des_vulsubtype,
                     des_cctype,
```

```
                    des_ccsubtype,
                    des_imptype,
                    des_impsubtype,
                    des_fin_sdg,
                    des_fin_jiaf,
                    EditDate,
                    category,
                    formduration)

names(rix)<-c("id",
              "user",
              "dataset_name",
              "org_type",
              "org",
              "url",
              "format_type",
              "format_ext",
              "pub_yr",
              "cca_themes",
              "cca_groups",
              "continent",
              "iso3c",
              "countries",
              "country_label",
              "spat_res_type",
              "spat_res",
              "start_yr",
              "end_yr",
              "haz_type",
              "haz_subtype",
              "haz_subsubtype",
              "exp_type",
              "exp_subtype",
              "vul_type",
              "vul_subtype",
              "cc_type",
              "cc_subtype",
              "imp_type",
              "imp_subtype",
              "sdg",
              "jiaf",
              "submit_date",
              "category",
              "form_duration"
              )

rix%<>%lapply(function(x) {if (is.factor(x)) x<-as.character(x) ; return(x)})%>%as.data.frame(stringsAs
```

Let's have a look! Firstly, let's have a look at the organisation types:

```
ggplot(rix)+geom_bar(aes(org_type,fill=org_type)) + ggtitle("RiX Organisation Type Bar Chart") +
  scale_fill_discrete(name="Organisation Type")+
  theme(axis.text.x=element_blank(),
```

```
        plot.title=element_text(hjust=0.5))
```



Now let's look at the different organisation names that we are using the most:

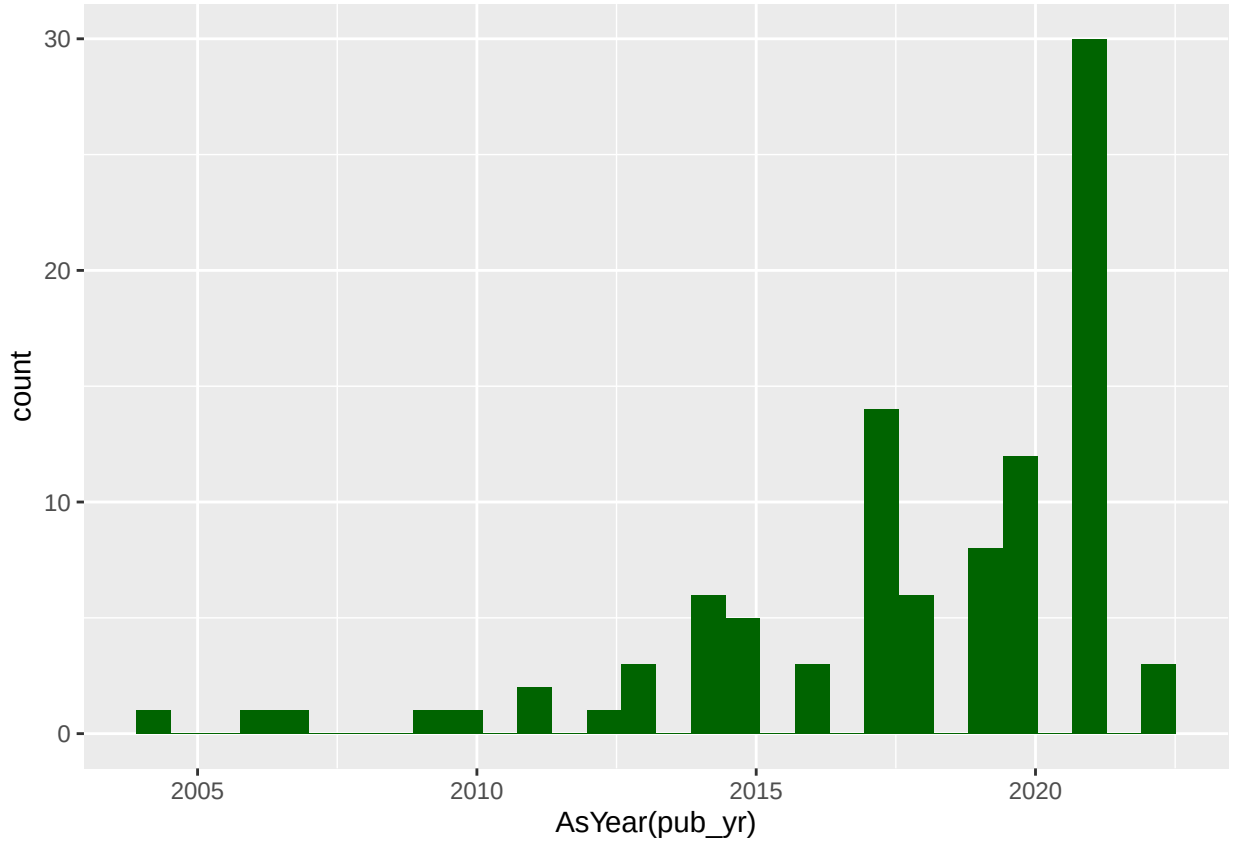
```
tmp<-rix%>%group_by(org)%>%summarise(Frequency=length(org))%>%arrange(desc(Frequency))
colnames(tmp)[1]<-"Organisation"
tmp
```

```
## # A tibble: 60 x 2
##    Organisation                                                Frequency
##    <chr>                                                           <int>
##  1 "UNEP (GRID)"                                                      13
##  2 "Water & Development Research Group, Aalto University"              6
##  3 "National Oceanic and Atmospheric Administration (NOAA)"            4
##  4 "World Bank"                                                        4
##  5 "Famine and Early Warning Systems Network (FEWSNET)"                3
##  6 "Resource Watch- World Resources Institute (WRI)"                   3
##  7 "World Bank "                                                       3
##  8 "World Health Organization (WHO)"                                   3
##  9 "WorldPop"                                                          3
## 10 "European Commission"                                               2
## # ... with 50 more rows
```

```
rm(tmp)
```

How about the year of publication?

```
ggplot(rix)+geom_histogram(aes(AsYear(pub_yr)),fill="darkgreen",bins=30)
```

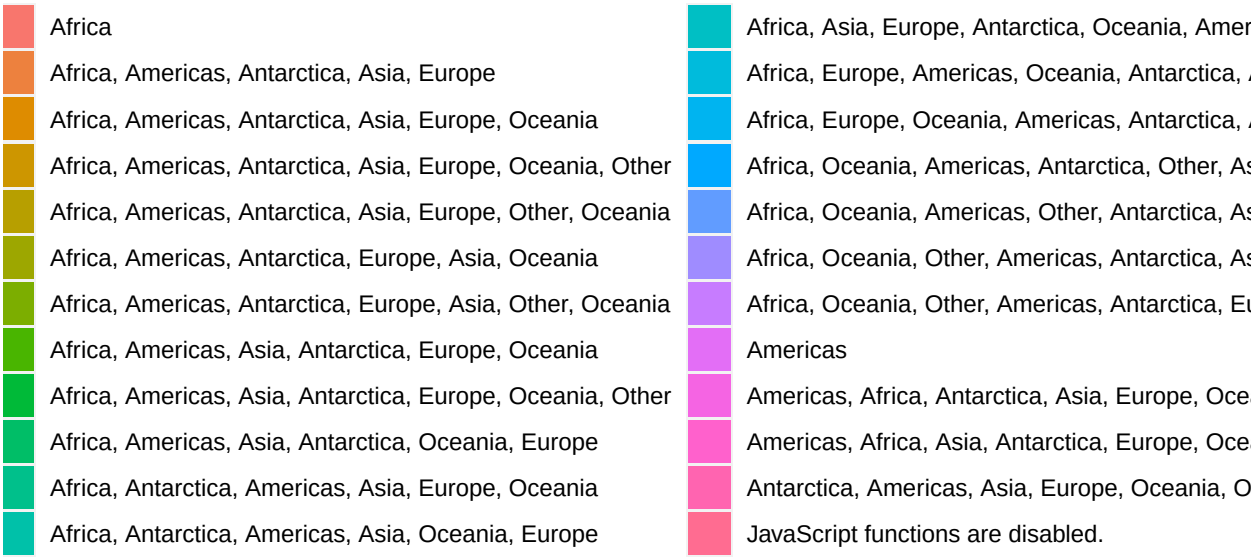


For the datasets that we are collecting, which continents are mostly covered so far?

```
ggplot(rix)+geom_bar(aes(continent,fill=continent))+ ggtitle("RiX Continents Bar Chart") +
  scale_fill_discrete(name="Continents")+
  theme(axis.text.x=element_blank(),
        plot.title=element_text(hjust=0.5))
```

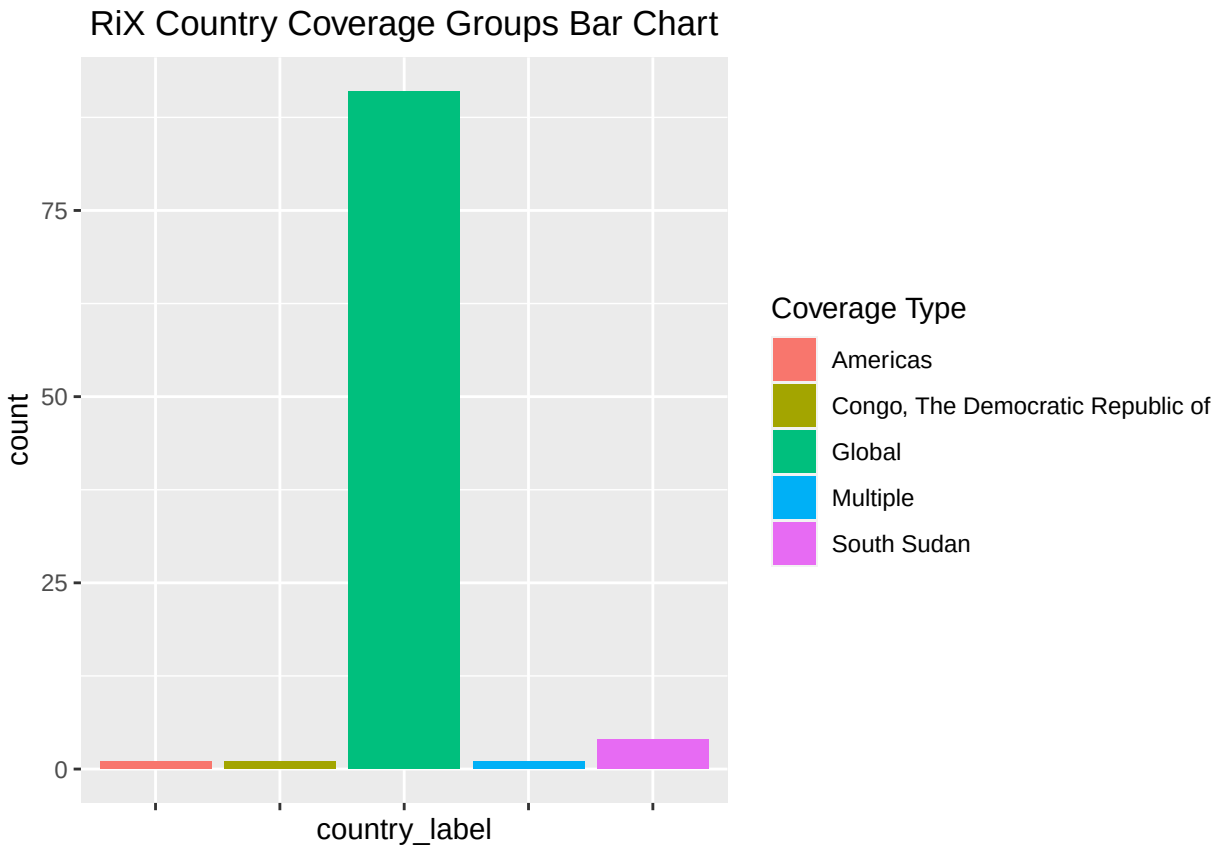
Bar Chart

Continents



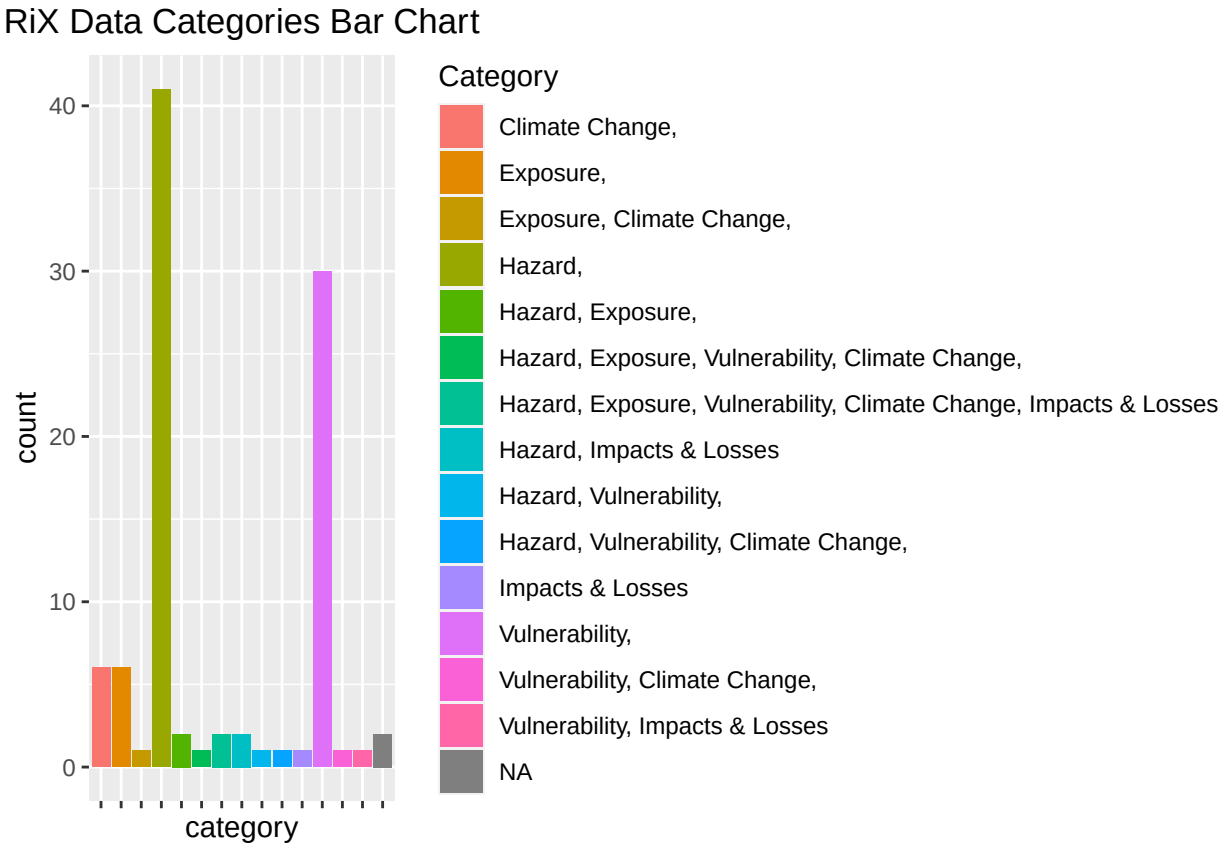
For the datasets that we are collecting, are they global, regional, multiple or single countries?

```
ggplot(rix)+geom_bar(aes(country_label,fill=country_label))+ ggtitle("RiX Country Coverage Groups Bar C
  scale_fill_discrete(name="Coverage Type")+
  theme(axis.text.x=element_blank(),
        plot.title=element_text(hjust=0.5))
```



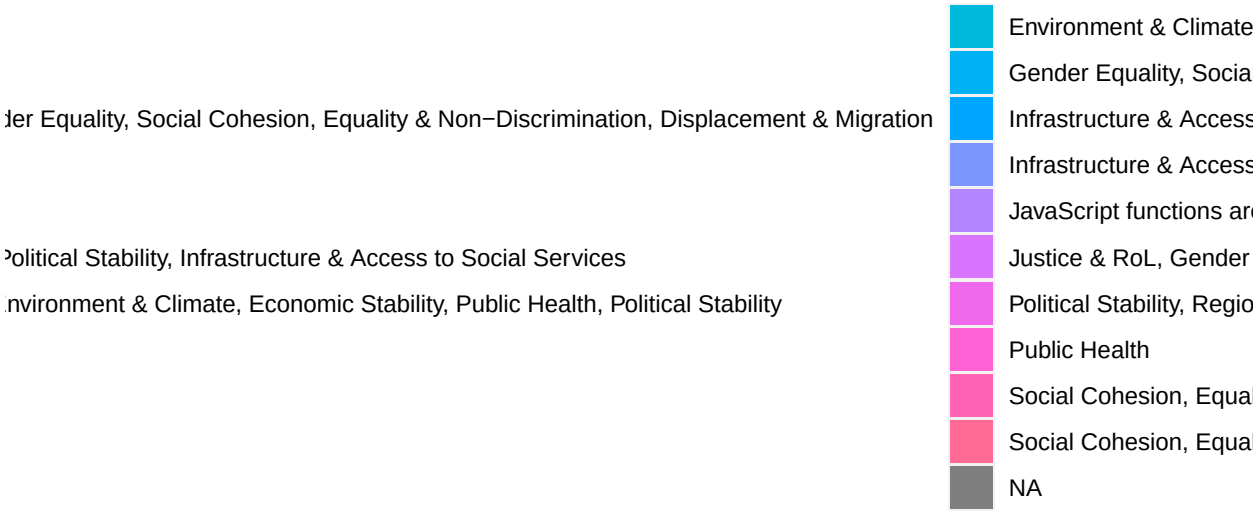
How about the categories that the datasets span?

```
ggplot(rix)+geom_bar(aes(category,fill=category)) + ggtitle("RiX Data Categories Bar Chart") +
  scale_fill_discrete(name="Category")+
  theme(axis.text.x=element_blank(),
        plot.title=element_text(hjust=0.5))
```



How about the 13 SDG risk areas?

```
ggplot(rix)+geom_bar(aes(sdg,fill=sdg)) + ggtitle("RiX 13 SDG Risk Areas Bar Chart") +
  scale_fill_discrete(name="SDG Risk Area")+
  theme(axis.text.x=element_blank(),
        plot.title=element_text(hjust=0.5))
```



How about the JIAF themes?

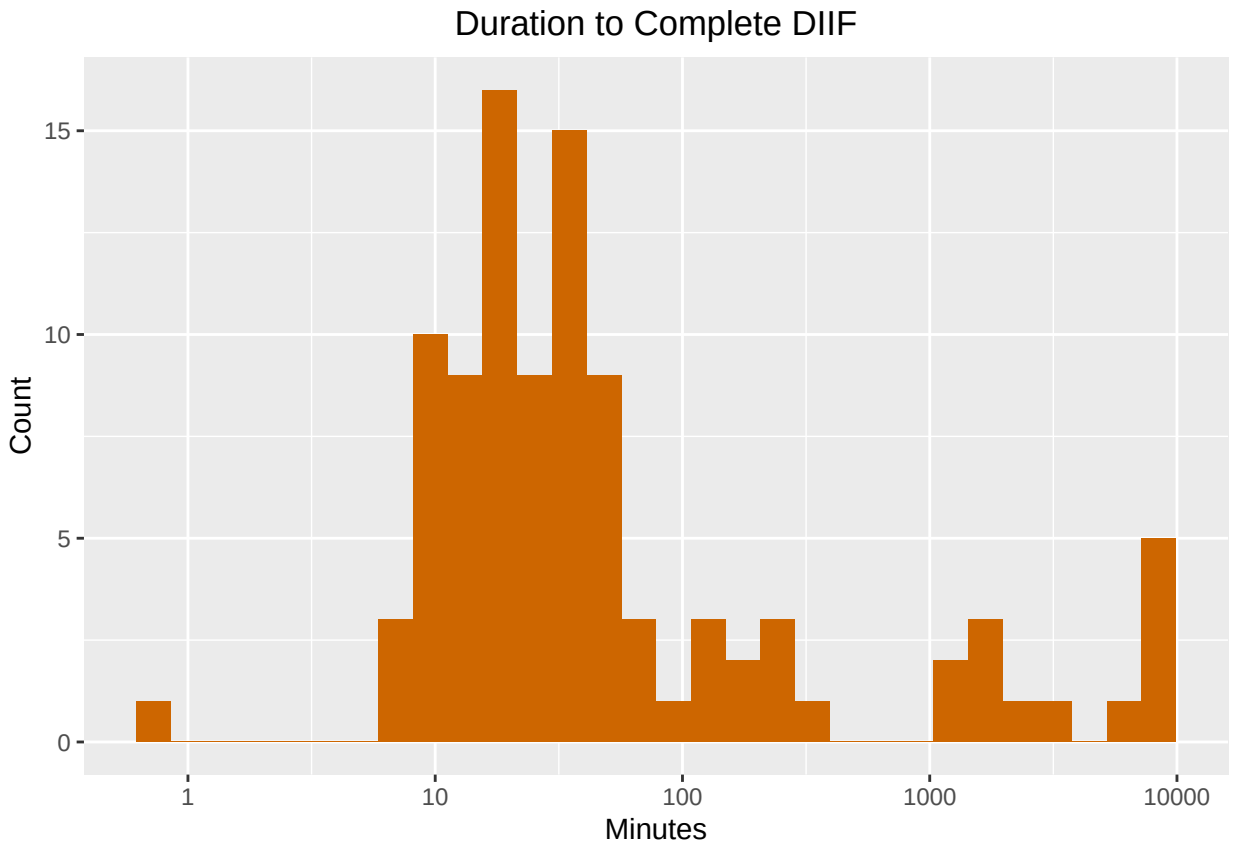
```
ggplot(rix)+geom_bar(aes(jiaf,fill=jiaf)) + ggtitle("RiX JIAF Themes Bar Chart") +
  scale_fill_discrete(name="JIAF Themes")+
```

```
  theme(axis.text.x=element_blank(),
        plot.title=element_text(hjust=0.5))
```

How long does it take to complete the form (in minutes), including having to resubmit?

```
ggplot(filter(rix,form_duration>0))+geom_histogram(aes(form_duration),fill="darkorange3")+
  scale_x_log10() + xlab("Minutes") + ylab("Count") + ggtitle("Duration to Complete DIIF")+
  theme(plot.title=element_text(hjust=0.5))
```

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


Duration to Complete DIIF
Count
15
10
5
0
1
10
100
1000
10000
Minutes