Mission2

Ujjval Sharma

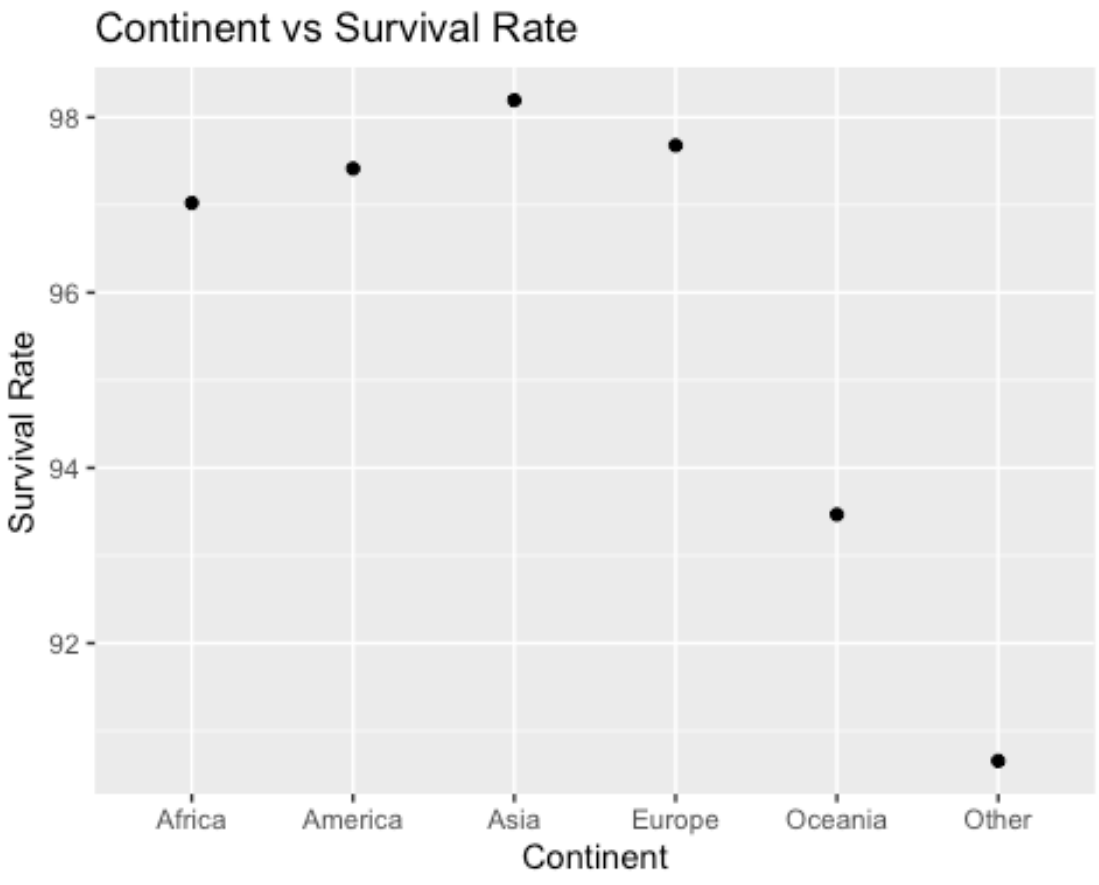
4/13/2021

R Markdown

This would tell the survial rate for different contients

```
library(readr)
library(tidyverse)

library(dplyr)
all_cases <- read_csv("/Users/ujjvalsharma/Documents/data-science/
project work/mission2/data-set/ddf--datapoints--cases--by--geo_id--
day.csv")

all_death <- read_csv("/Users/ujjvalsharma/Documents/data-science/
project work/mission2/data-set/ddf--datapoints--deaths--by--geo_id--
day.csv")

all_geo_names <- read_csv("/Users/ujjvalsharma/Documents/data-science/
project work/mission2/data-set/ddf--entities--geo_id.csv")

all_cases[, 3] =all_cases[, 3] + 1 #adding +1 for 0 cases
all_death[, 3] =all_death[, 3] + 1 #adding +1 for 0 death

# this code is used to get survivial rate for each contient
all_cases_grp_geo<-all_cases %>%
  group_by(geo_id) %>%
  summarize(sum(cases, na.rm=TRUE))

all_death_grp_geo<-all_death %>%
  group_by(geo_id) %>%
  summarize(sum(deaths, na.rm=TRUE))

names(all_cases_grp_geo)[2] <- "cases"
names(all_death_grp_geo)[2] <- "deaths"


all_cases_grp_contient<-inner_join(all_cases_grp_geo, all_geo_names,
```

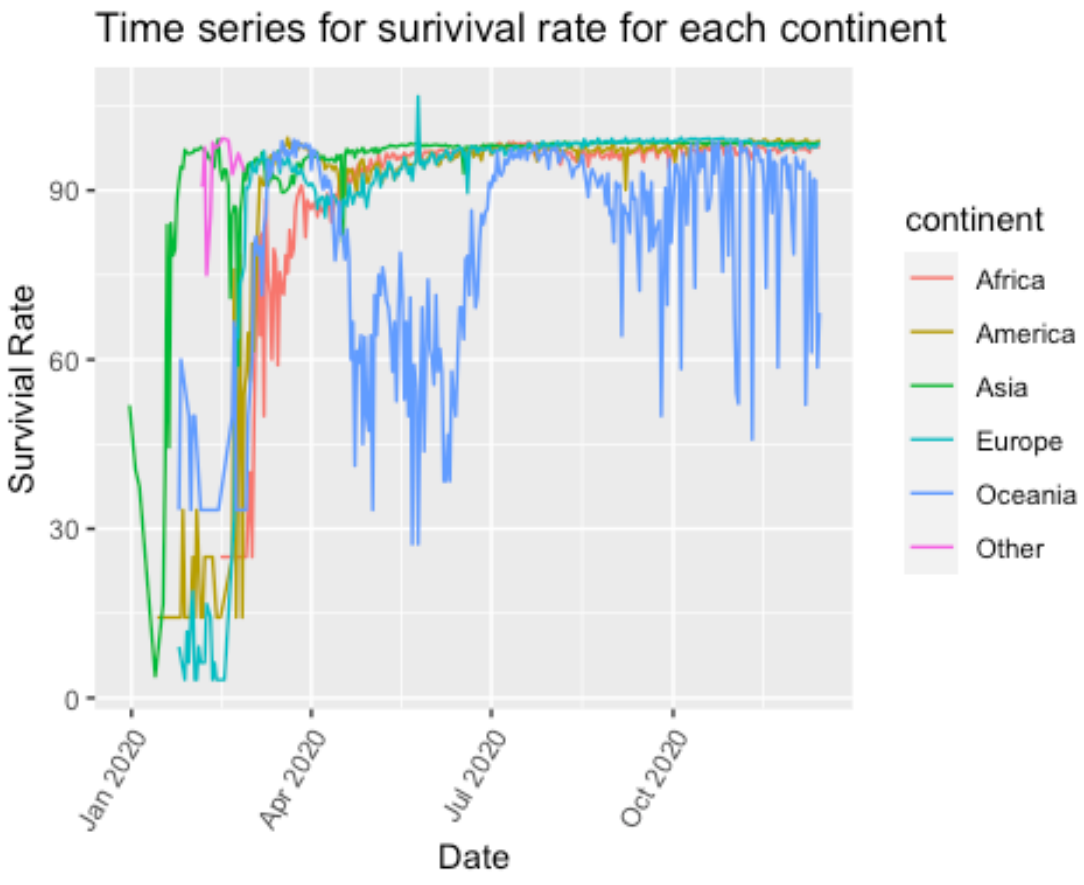
```r
by=c("geo_id"="geo_id"))  %>%
  group_by(continent_exp) %>%
  summarize(sum(cases, na.rm=TRUE))

all_death_grp_contient<-inner_join(all_death_grp_geo, all_geo_names,
by=c("geo_id"="geo_id"))  %>%
  group_by(continent_exp) %>%
  summarize(sum(deaths, na.rm=TRUE))

names(all_cases_grp_contient)[2] <- "cases"
names(all_death_grp_contient)[2] <- "deaths"

cases_death_contient<- inner_join(all_cases_grp_contient,
all_death_grp_contient, by=c("continent_exp"="continent_exp"))


cases_death_contient <- transform(cases_death_contient,
survival_count=cases-deaths)


cases_death_contient<-cases_death_contient %>%
filter(survival_count>0)

cases_death_contient <- transform(cases_death_contient,
survival_rate=(survival_count/cases)*100)


ggplot(cases_death_contient, aes(x = continent_exp,y = survival_rate))
+
  geom_point()+
  labs(title="Continent vs Survival Rate",
       x ="Continent", y = "Survival Rate")
```



This would tell the time series for surivival rate.

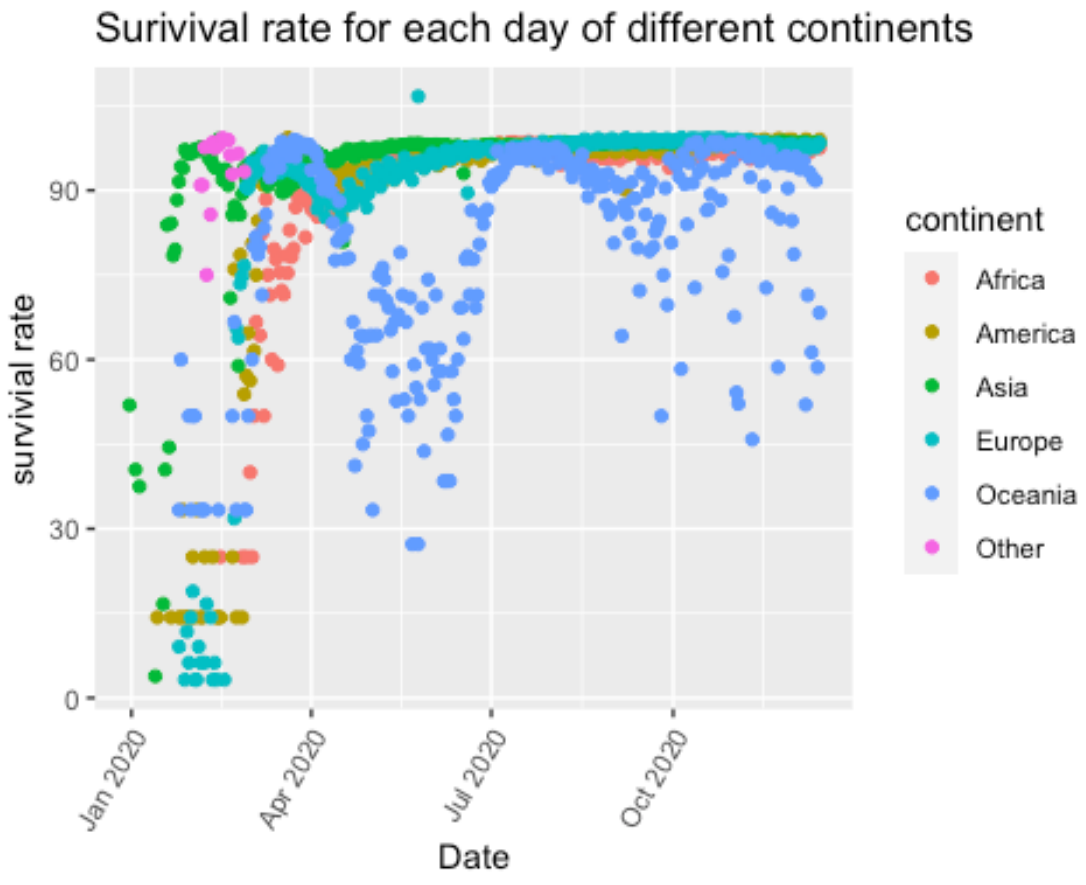
```
all_day_cases_grp_contient<-inner_join(all_cases, all_geo_names,
by=c("geo_id"="geo_id"))  %>%
  group_by(continent_exp,day) %>%
  summarize(sum(cases, na.rm=TRUE))

## `summarise()` has grouped output by 'continent_exp'. You can
override using the `.groups` argument.

all_day_death_grp_contient<-inner_join(all_death, all_geo_names,
by=c("geo_id"="geo_id"))  %>%
  group_by(continent_exp,day) %>%
  summarize(sum(deaths, na.rm=TRUE))

## `summarise()` has grouped output by 'continent_exp'. You can
override using the `.groups` argument.
```

```
names(all_day_cases_grp_contient)[3] <- "cases"
names(all_day_death_grp_contient)[3] <- "deaths"


all_cases_death_day<-inner_join(all_day_death_grp_contient,
all_day_cases_grp_contient,
by=c("continent_exp"="continent_exp","day"="day"))


all_cases_death_day <- transform(all_cases_death_day,
survival_count=cases-deaths)

all_cases_death_day<-all_cases_death_day %>% filter(survival_count>0)

all_cases_death_day <- transform(all_cases_death_day, survival_rate=
ifelse(cases>0,(survival_count/cases)*100,100))

all_cases_death_day <- transform(all_cases_death_day, covid_date=
as.Date(paste(paste(substr(day, 1, 4),substr(day, 5, 6), sep="-"),

substr(day, 7, 8),sep="-")))


names(all_cases_death_day)[names(all_cases_death_day) ==
"continent_exp"] <- "continent"

  ggplot(all_cases_death_day, aes(x=covid_date,
y=survival_rate,color=continent)) +
  geom_line() +
  theme(axis.text.x=element_text(angle=60, hjust=1))+
    labs(title="Time series for surivival rate for each continent",
         x ="Date", y = "Survivial Rate")
```



This would tell the survial rate for different contients for each day.

```
  ggplot(all_cases_death_day, aes(x=covid_date,
y=survival_rate,color=continent)) +
    geom_point() +
    theme(axis.text.x=element_text(angle=60, hjust=1))+
    labs(title="Surivival rate for each day of different continents",
         x ="Date", y = "survivial rate")
```


Surivival rate for each day of different continents
survivial rate
90
60
30
0
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Date
continent
Africa
America
Asia
Europe
Oceania
Other

```
```