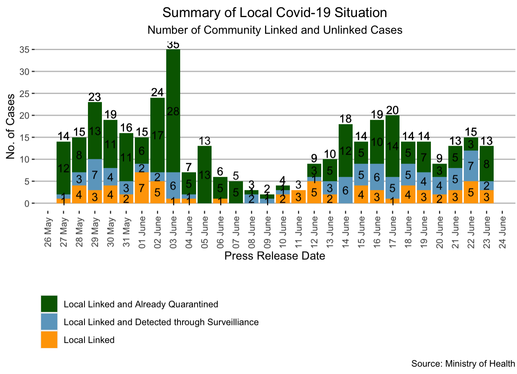
Tut9

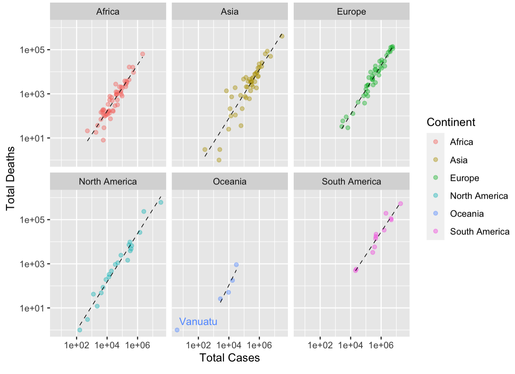
Ananya_Gupta

10/16/2021

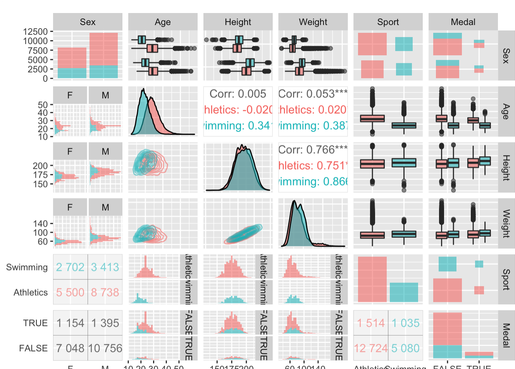
```
#Qn 1
covid<-read_csv("../data/covid_sg.csv")
covid$Date<-as.Date(covid$Date,"%d/%m/%Y")

covid<-covid %>%
  mutate(new_date=format(Date,format="%d %B"))%>%
  pivot_longer(2:4, names_to = "type", values_to = "cases")%>%
  arrange(Date,rev(type))%>%
  group_by(Date)%>%
  mutate(label_y=cumsum(cases)- (0.5 * cases), tot=sum(cases))%>%
  filter(cases>0)

covid %>%
  ggplot()+
  geom_col(mapping=aes(x=Date, y=cases, fill=type))+
  scale_fill_manual(values=c("#006400", "#6CA6C7", "#FFA500"),
                    name="",
                    labels=c("Local Linked and Already Quarantined",
                             "Local Linked and Detected through Surveilliance",
                             "Local Linked"))+
  theme(legend.position = "bottom",
        legend.justification = "left",
        legend.direction= "vertical")+
  theme(axis.text.x=element_text(angle=90, hjust=0.5,vjust=0.5))+
  labs(x="Press Release Date",
       y="No. of Cases",
       title="Summary of Local Covid-19 Situation",
       caption="Source: Ministry of Health",
       subtitle = "Number of Community Linked and Unlinked Cases")+
  geom_text(aes(x=Date,y=label_y,label=cases))+
  geom_text(aes(x=Date,y=tot,label=tot), vjust=-0.2)+
  theme(panel.grid.major.x = element_blank(),
        panel.grid.major.y = element_line(linetype = "solid",
                                          color="grey"),
        panel.background = element_blank(),
        plot.title=element_text(hjust=0.5),
        plot.subtitle=element_text(hjust=0.5))+
  scale_y_continuous(breaks = seq(0, 35, by= 5))+
  scale_x_date(breaks=seq(as.Date("2021-05-26"), as.Date("2021-06-24"), by = "1 day"),
               date_labels = "%d %B")
```




```
#Qn 2
covid_world<-read_csv("../data/covid_world_today.csv")

fil<-covid_world %>%
  filter(location!="Vanuatu")

point<-covid_world %>%
  filter(location=="Vanuatu")

ggplot()+
  geom_jitter(covid_world,mapping=aes(x=total_cases, y=total_deaths, color=continent), alpha=0.4)+
  geom_smooth(fil,mapping=aes(x=total_cases,y=total_deaths,color=continent),linetype="dashed",color="black",size=
0.3,
              method="lm",se=FALSE)+
  scale_x_log10()+
  scale_y_log10()+
  facet_wrap(~continent)+
  labs(x="Total Cases", y="Total Deaths")+
  scale_color_discrete(name ="Continent")+
  geom_text(data=point,mapping=aes(x=total_cases, y=total_deaths,label=location, color=continent), nudge_x = 1.4,
nudge_y=0.4, size=3.5, show.legend=FALSE)
```




```
#Qn 3
noc <- read.csv("../data/athlete_events.csv")

noc %>%
  filter(Year >= 1970, Season=="Summer", Sport %in% c("Athletics", "Swimming")) %>%
  group_by(ID) %>%
  summarise(Sex=Sex[1], Age=min(Age, na.rm=FALSE),
            Height = mean(Height, na.rm=FALSE),
            Weight=mean(Weight, na.rm=FALSE), Sport=Sport[1], Medal=any(!is.na(Medal)),
            .groups="drop") -> noc_tmp

ggpairs(noc_tmp, mapping=aes(color=Sport, alpha=0.8), columns=2:7,
        lower=list(continuous="density", discrete="crosstable"),
        showStrips=TRUE)
```



- Observations
- More people in athletics than swimming so maybe a factor of strength or simply interest based
- More males in sports(athletics and swimming) overall than females, leading factor being different requirements from men and women maybe or different physique factor considerations for every sport
- Largely swimmers belong to lower age group than athletes. Majority of female and male swimmers belong to age group of 15-24 while athletes largely belong to age group of 20-35 yrs, so exploring the connection of age to the performance can help understand the reason better
- Most swimmers have higher weight and height which maybe interlinked since taller height relates to greater bone weight affecting the weight.
- Moreover, many outliers seen among athletes for their weight on the higher side and outliers in swimmers were more for their height on the lower side, highlighting important considerations for each sport