

RSV Summary = 46102, 8

```
head(data) %>%
  head(., n=10) %>%
  knitr::kable(., booktabs=T)
```

date	genbank	strain	host	country	genotype	species	length
NA	AB175817	NG_065_02	Unknown	Japan	NA	Human_orthopneumovirus	270
NA	AB485783	RSvi/Okinawa.JPN/209.08	Unknown	Japan	NA	Human_orthopneumovirus	387
2008-XX-XX	AB574207	HA08_80016	Human	Japan	NA	Human_orthopneumovirus	346
2005-12-13	AB603470	NG_108_05	Human	Japan	NA	Human_orthopneumovirus	330
2005-11-28	AB603475	NG_042_05	Human	Japan	NA	Human_orthopneumovirus	330
2010-01-19	AB683194	RSV/Kumamoto/9928/2010	Unknown	Japan	NA	Human_orthopneumovirus	393

```
data %>%
  group_by(species) %>%
  summarize(n=n()) %>%
  head(., n=10) %>%
  knitr::kable(., booktabs=T)
```

species	n
Avian_metapneumovirus	565
Avian_metapneumovirus;Human_metapneumovirus	20
Bovine_orthopneumovirus	518
Bovine_orthopneumovirus;Human_metapneumovirus	72
Bovine_orthopneumovirus;Human_orthopneumovirus	2
Canine_pneumovirus	27
Canine_pneumovirus;Murine_orthopneumovirus	2
Feline_pneumovirus	2
Gull_metapneumovirus	5
Human_metapneumovirus	9937

There're problematic strain names like "B1"

```
problematic = c("B1", "A")
data %>%
  subset(strain %in% problematic) %>%
  head(.) %>%
  mutate(
    genbank=substr(genbank, 0,20),
    date=substr(date, 0,20)
  ) %>%
  knitr::kable(., booktabs=T)
```

date	genbank	strain	host	country	genotype	species
NA	AF013255;GQ872026;AF	B1	Unknown	NA	NA	Human_orthopneumovirus
2011-12-29;2006-XX-X	DQ885231;KP853084;JQ	A	Human;Unknown	USA;Sudan;Brazil;Latvia	NA	Human_orthopneumovirus
2011-12-29;2006-XX-X	DQ885231;KP853084;JQ	A	Human;Unknown	USA;Sudan;Brazil;Latvia	NA	Human_orthopneumovirus
2011-12-29;2006-XX-X	DQ885231;KP853084;JQ	A	Human;Unknown	USA;Sudan;Brazil;Latvia	NA	Human_orthopneumovirus
2011-12-29;2006-XX-X	DQ885231;KP853084;JQ	A	Human;Unknown	USA;Sudan;Brazil;Latvia	NA	Human_orthopneumovirus
2011-12-29;2006-XX-X	DQ885231;KP853084;JQ	A	Human;Unknown	USA;Sudan;Brazil;Latvia	NA	Human_orthopneumovirus

1 genotypes

Placeholder for later

```
data %>%
  mutate(
    year=substr(date, 0,4),
    genotype = case_when(grepl("RSV_A", genotype)~ "A",
                          grepl("RSV_B", genotype)~ "B",
                          1==1 ~ genotype)
  ) %>%
```

```
subset(!is.na(year)) %>%  
ggplot(., aes(x=year, fill=genotype)) +  
geom_bar() +  
theme_bw() +  
theme(axis.text.x = element_text(angle=45))+  
labs(title="genotypes")
```

