TSAR: Clinical EDA

Joshua Burkhart

August 15, 2016

Scan for Missing Data

```
sum(is.na(train_symptom_df)) #0
```

[1] 0

```
sum(is.na(train_clinicl_df)) #4318
```

[1] 4318

```
sum(is.na(phas1_clinicl_df)) #60
```

[1] 60

```
sum(is.na(phas2_clinicl_df)) #88
```

[1] 88

```
sum(is.na(phas3_clinicl_df)) #62
```

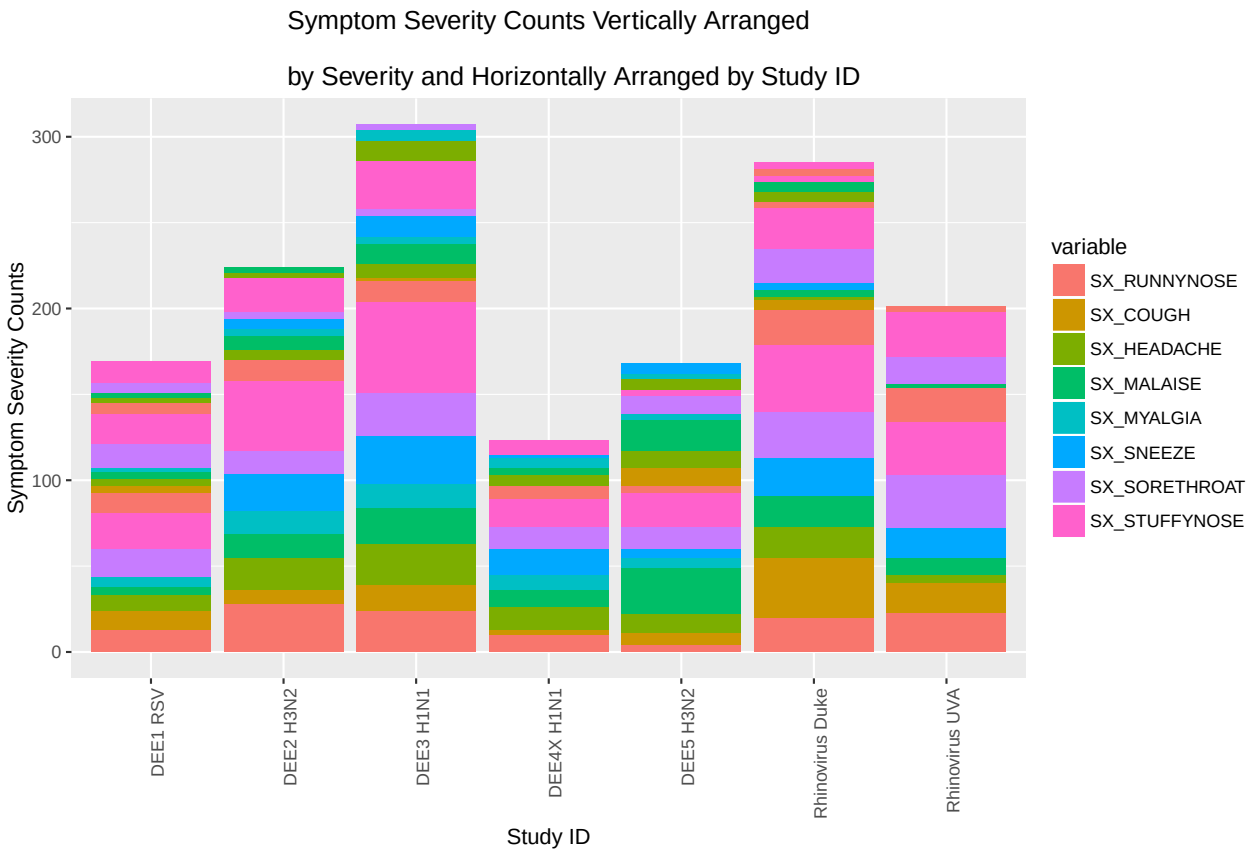
[1] 62

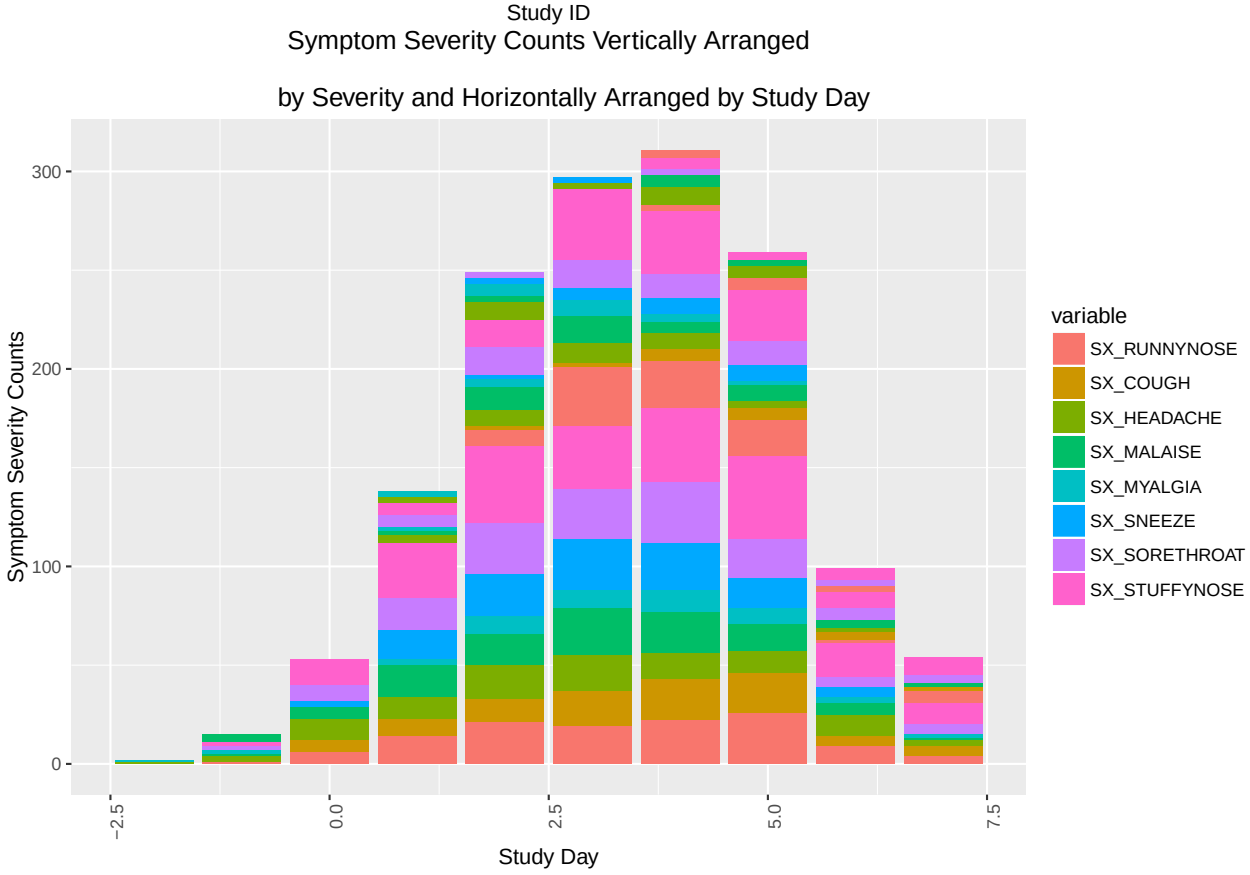
Fix Missing Data

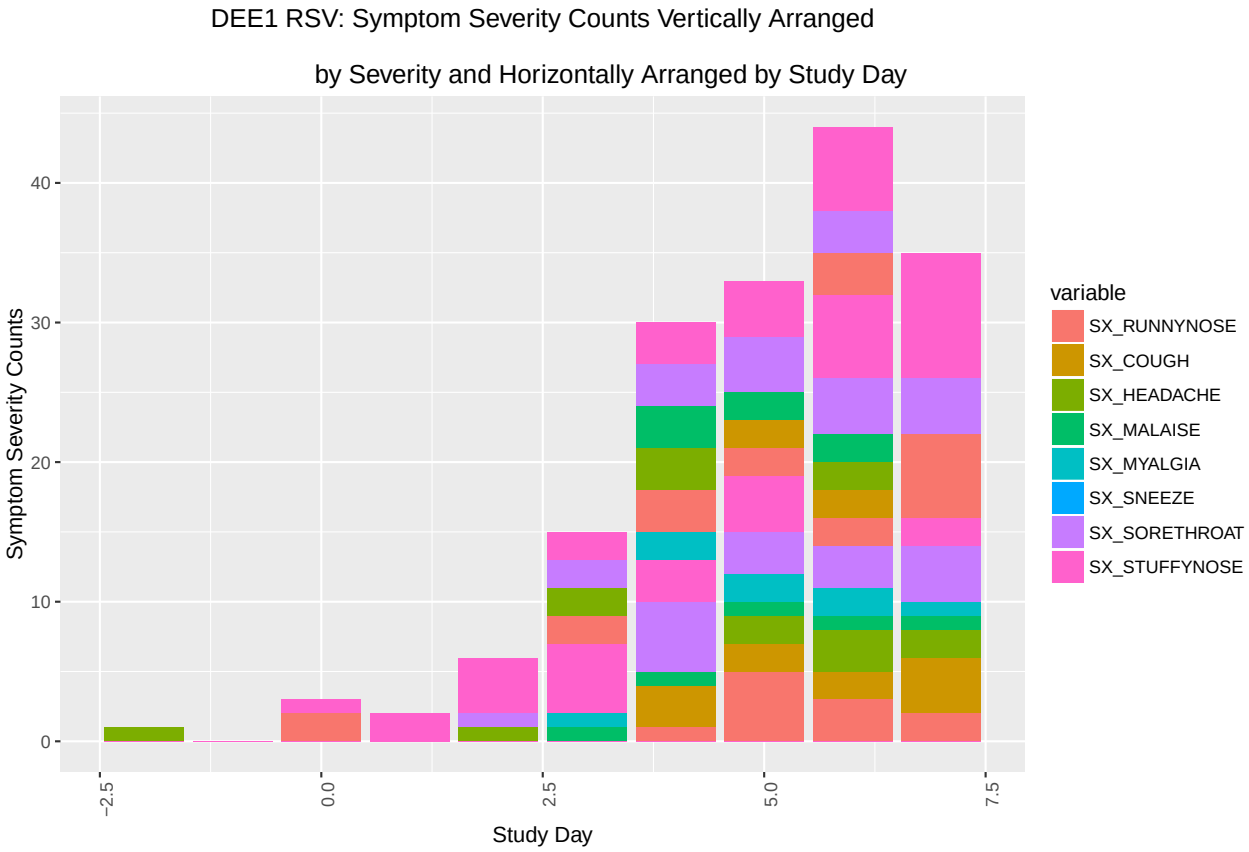
All NA's are found in EARLYTX and SHAM. Change those columns to binary values.

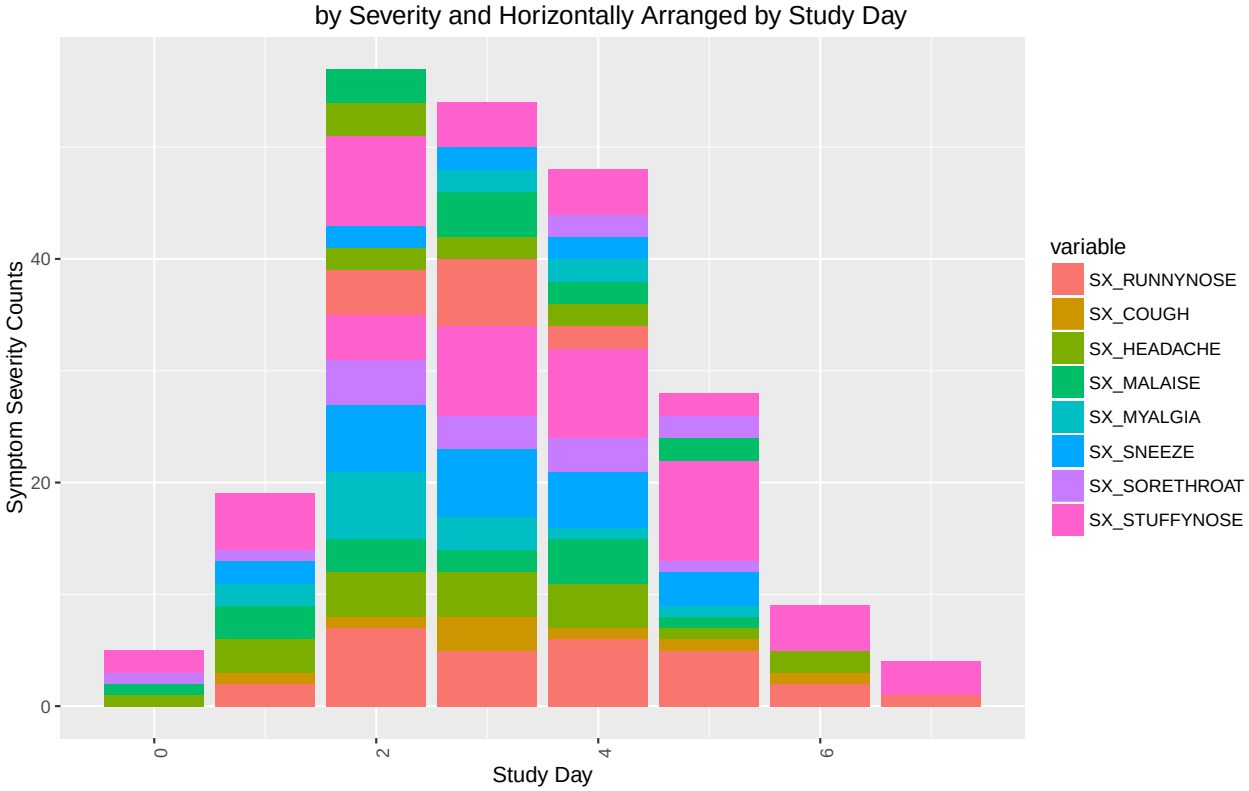
Checking Distributions

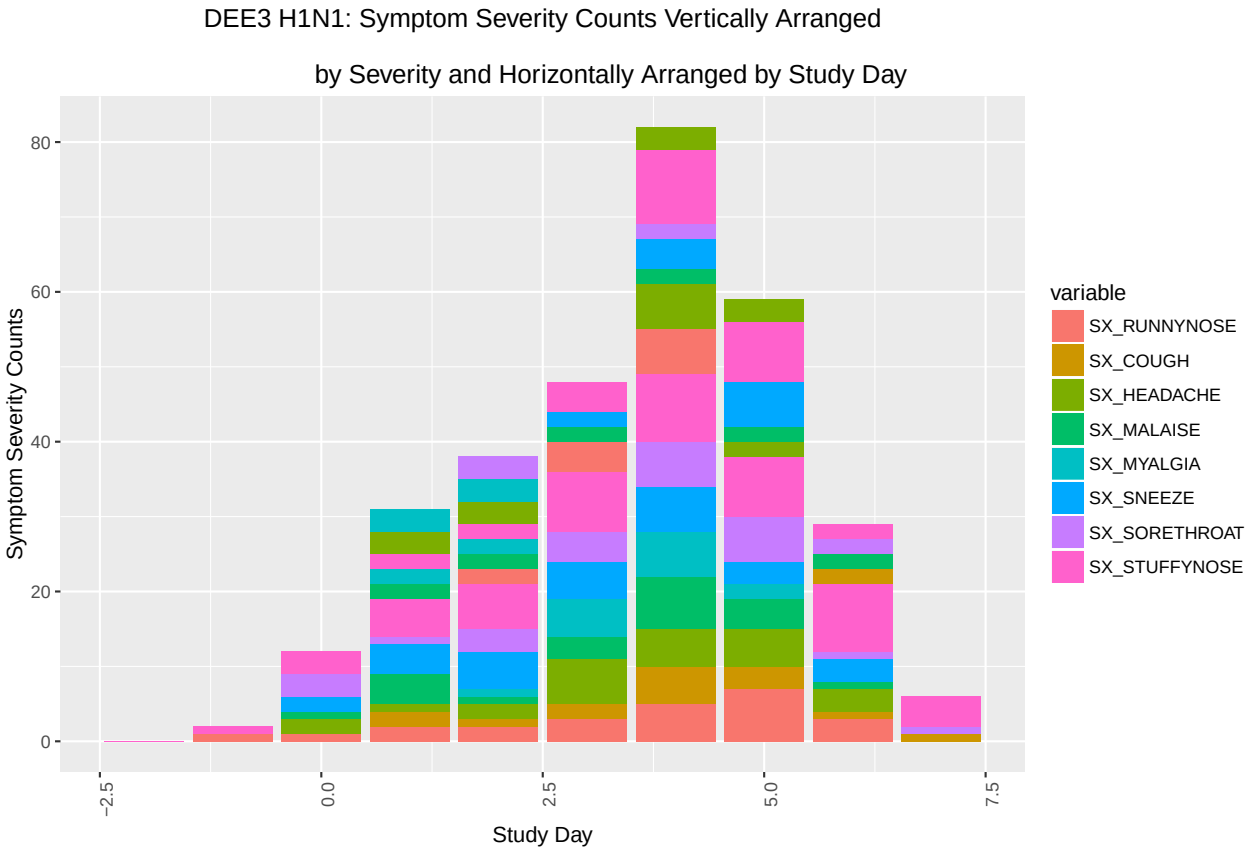
Symptoms

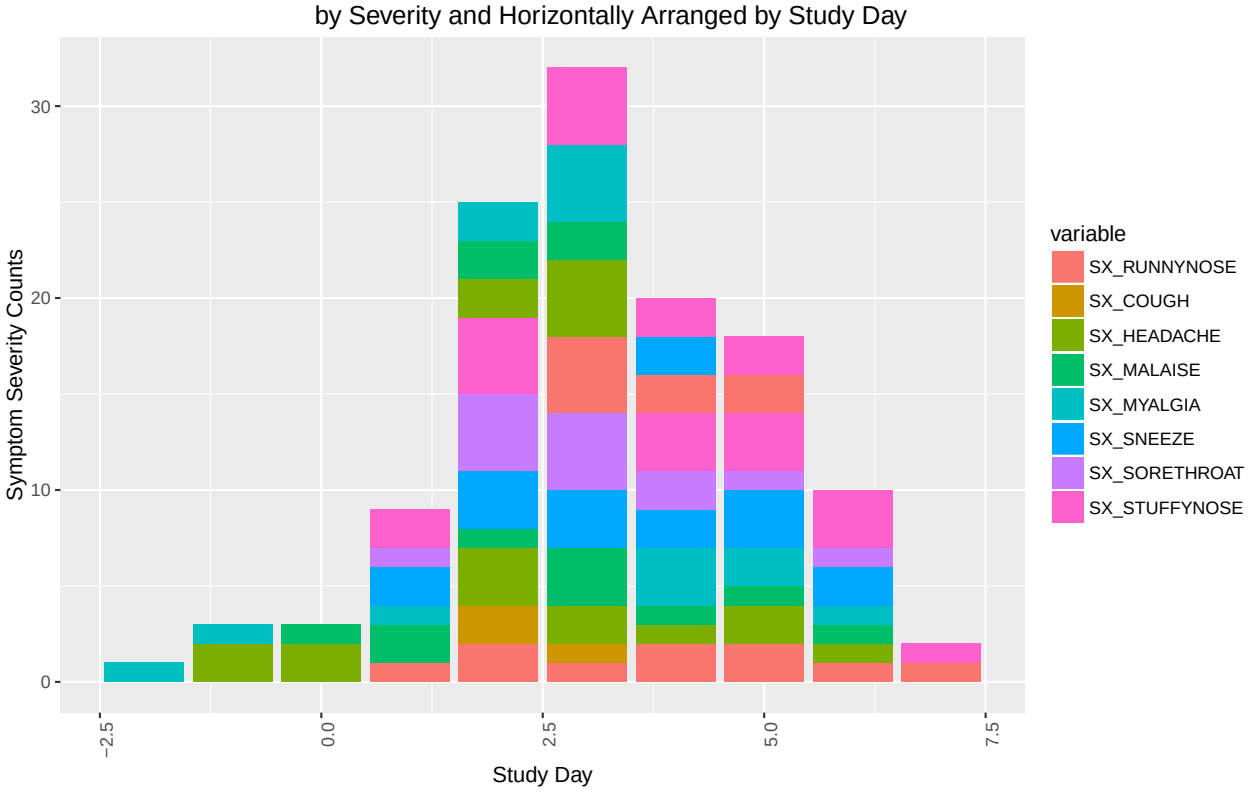

Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study ID
Symptom Severity Counts
300
200
100
0
DEE1 RSV
DEE2 H3N2
DEE3 H1N1
DEE4X H1N1
DEE5 H3N2
Rhinovirus Duke
Rhinovirus UVA
Study ID
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

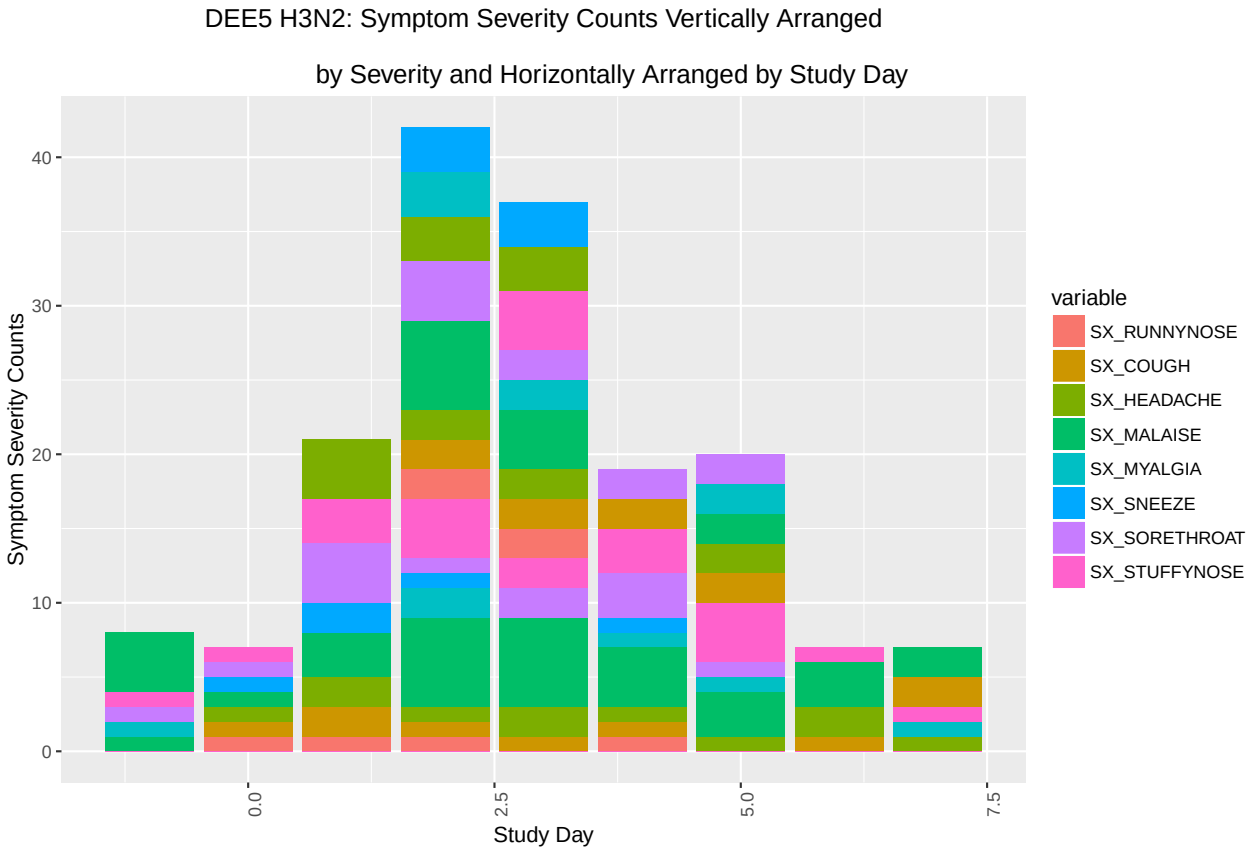

Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
300
200
100
0
−2.5
0.0
2.5
5.0
7.5
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

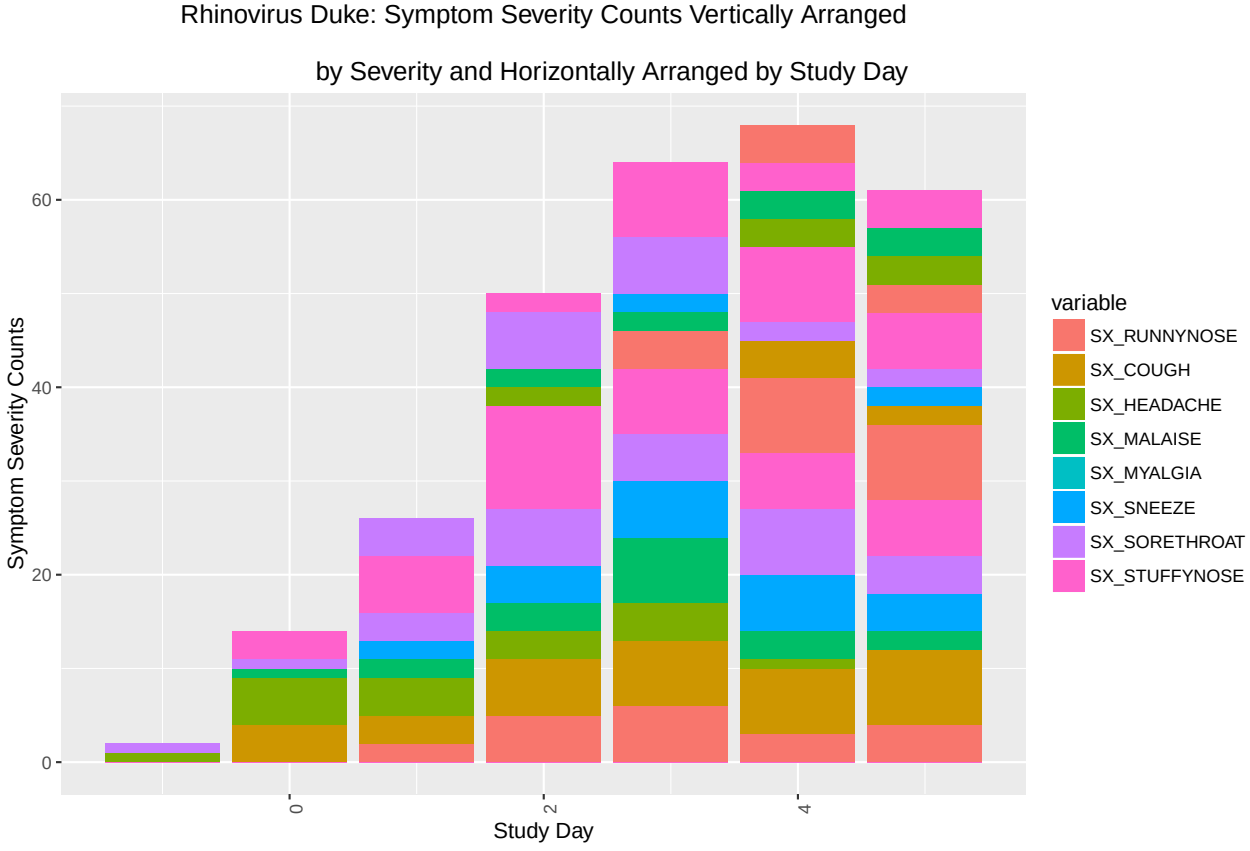

DEE1 RSV: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
0
10
20
30
40
−2.5
0.0
2.5
5.0
7.5
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

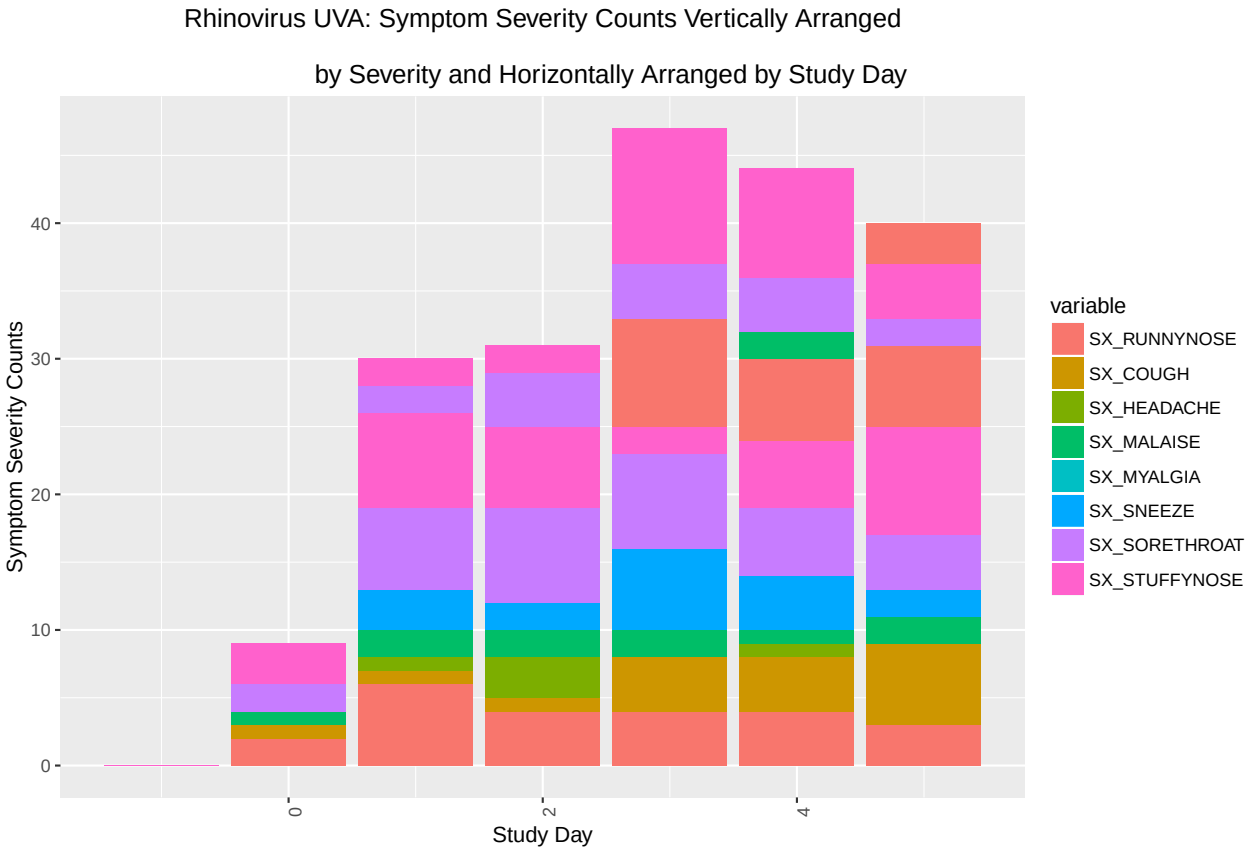

DEE2 H3N2: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
0
20
40
0
2
4
6
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

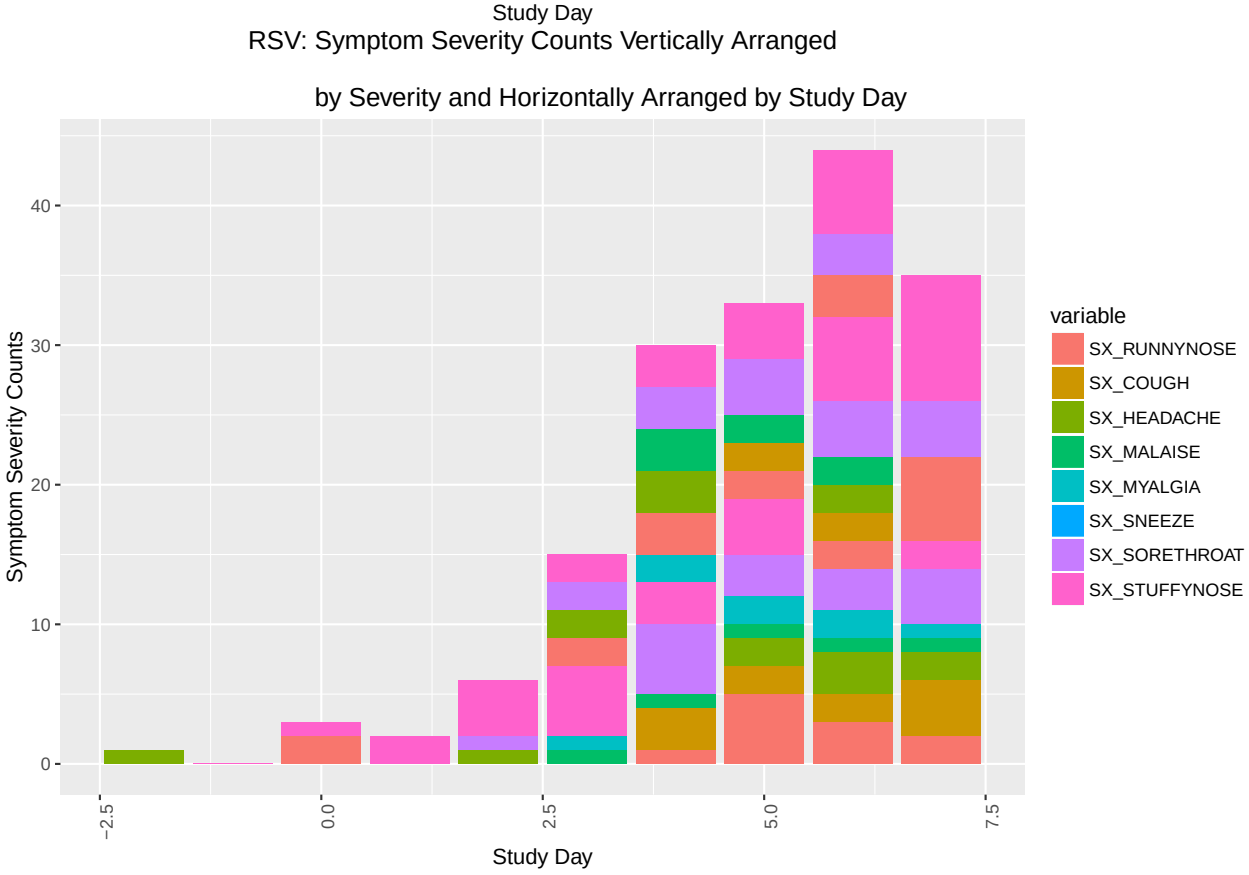

DEE3 H1N1: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
0
20
40
60
80
−2.5
0.0
2.5
5.0
7.5
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

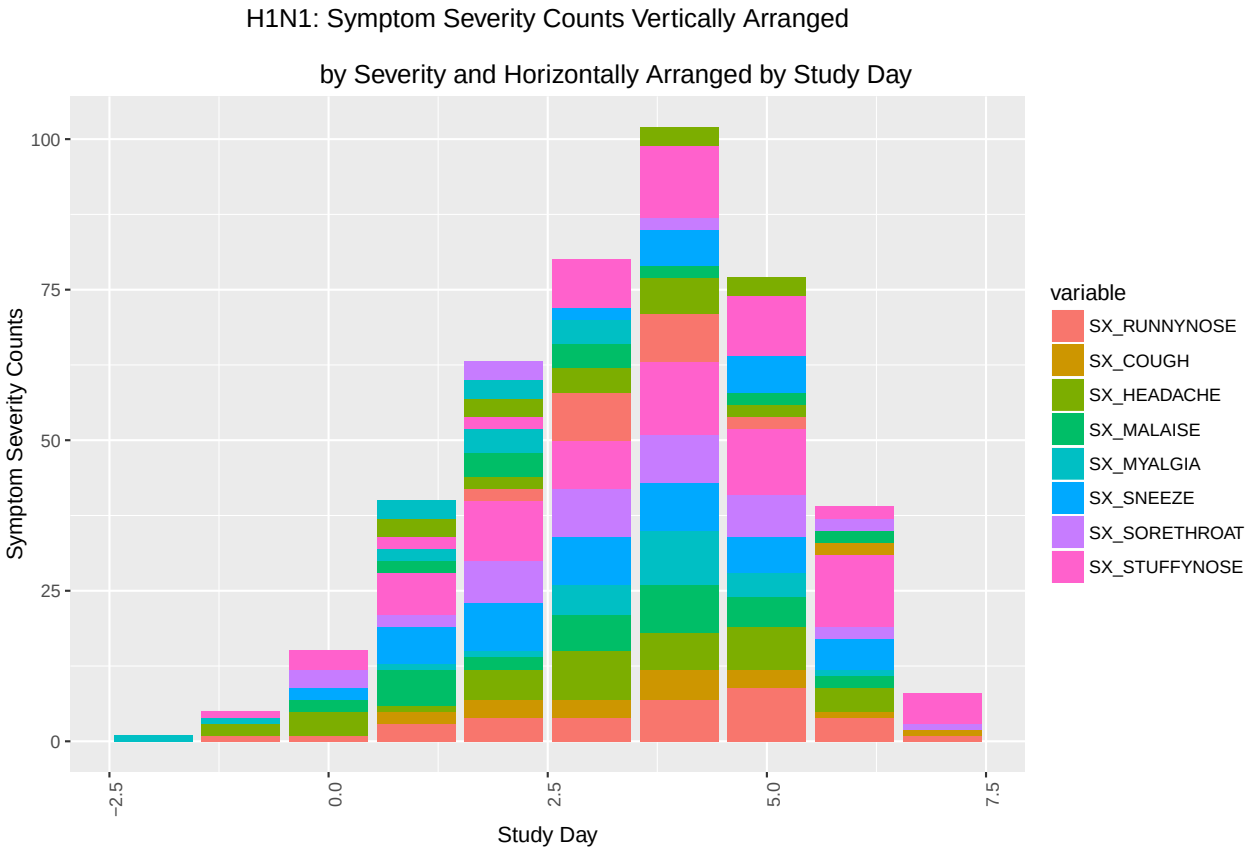

DEE4X H1N1: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
0
10
20
30
−2.5
0.0
2.5
5.0
7.5
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

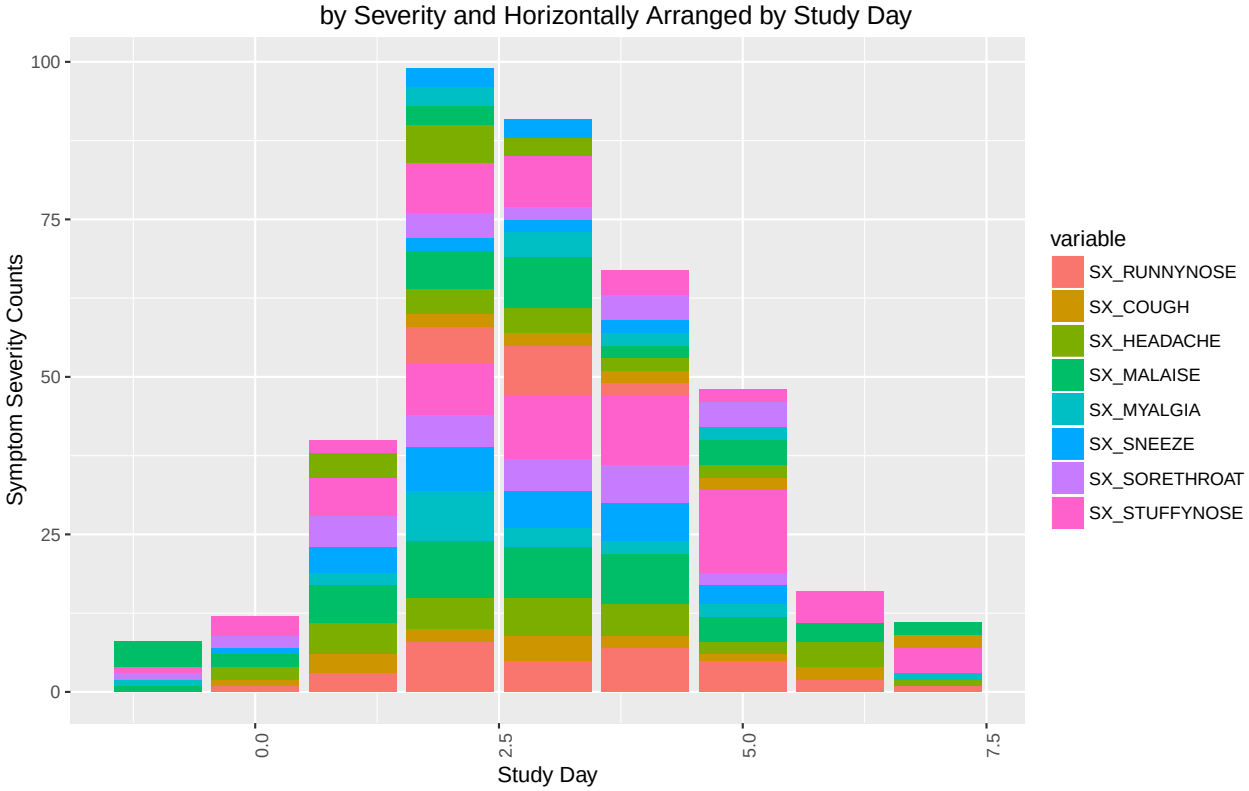

DEE5 H3N2: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
0
10
20
30
40
0.0
2.5
5.0
7.5
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

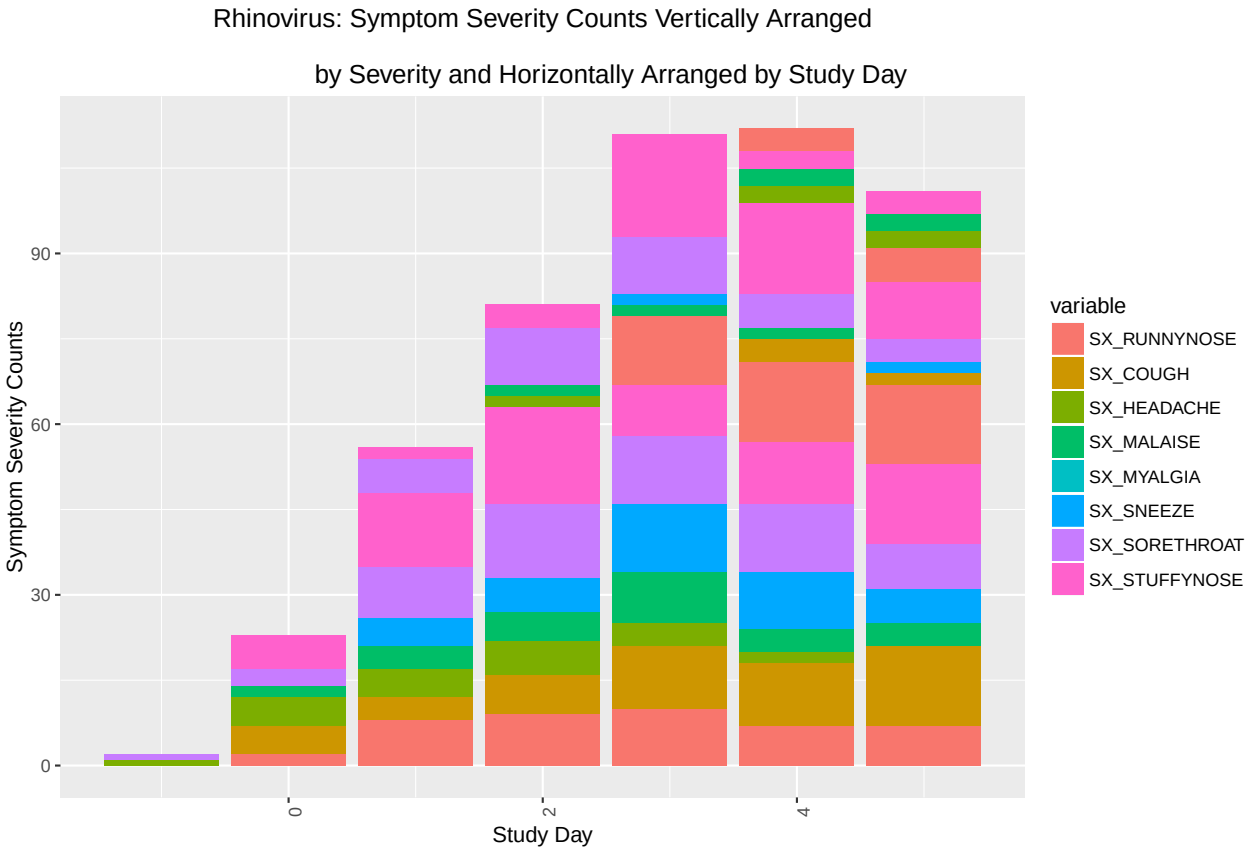

Rhinovirus Duke: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
0
20
40
60
0
2
4
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE


Rhinovirus UVA: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
0
10
20
30
40
0
2
4
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE


RSV: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
0
10
20
30
40
−2.5
0.0
2.5
5.0
7.5
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

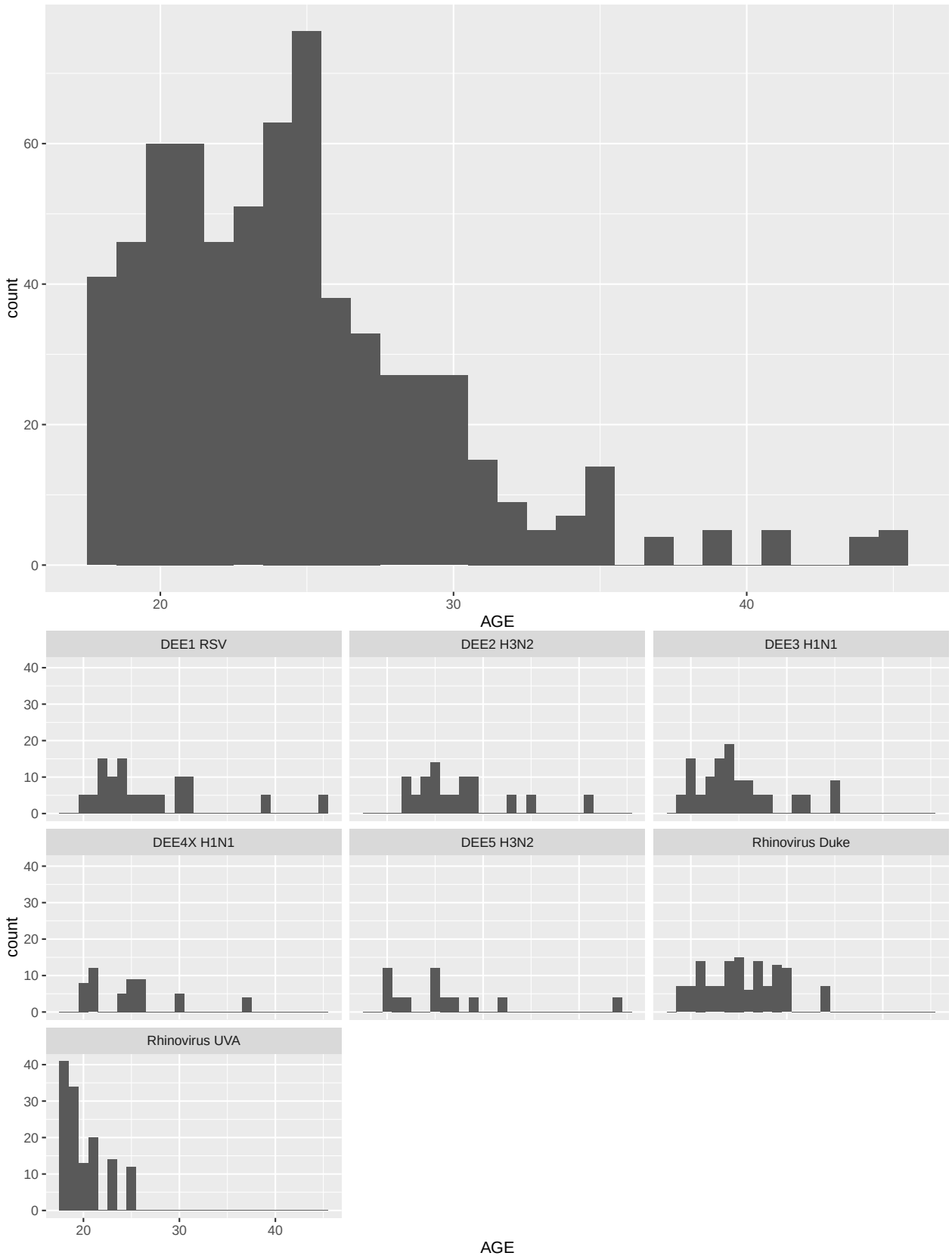

H1N1: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
100
75
50
25
0
−2.5
0.0
2.5
5.0
7.5
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

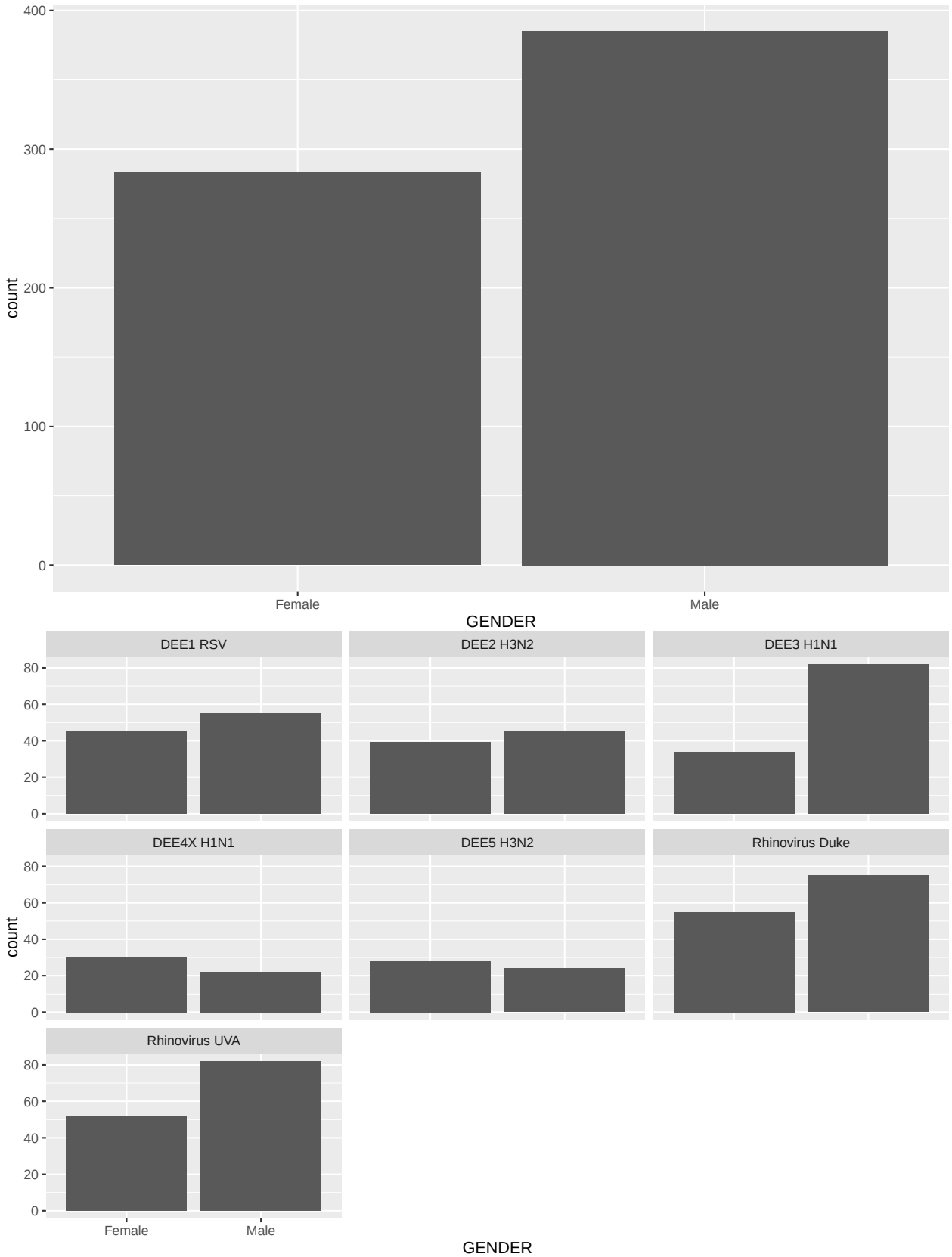

H3N2: Symptom Severity Counts Vertically Arranged
by Severity and Horizontally Arranged by Study Day
Symptom Severity Counts
100
75
50
25
0
0.0
2.5
5.0
7.5
Study Day
variable
SX_RUNNYNOSE
SX_COUGH
SX_HEADACHE
SX_MALAISE
SX_MYALGIA
SX_SNEEZE
SX_SORETHROAT
SX_STUFFYNOSE

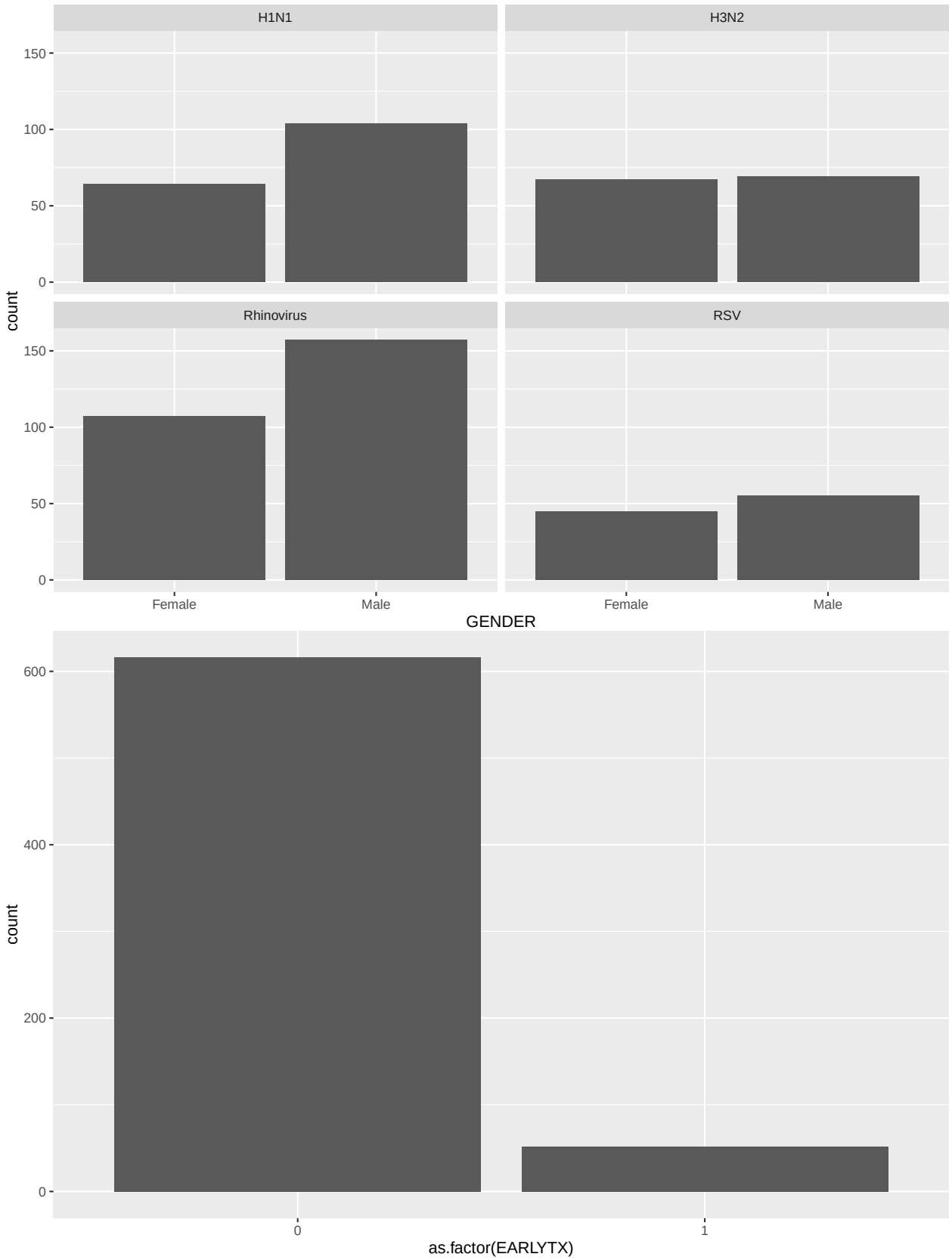


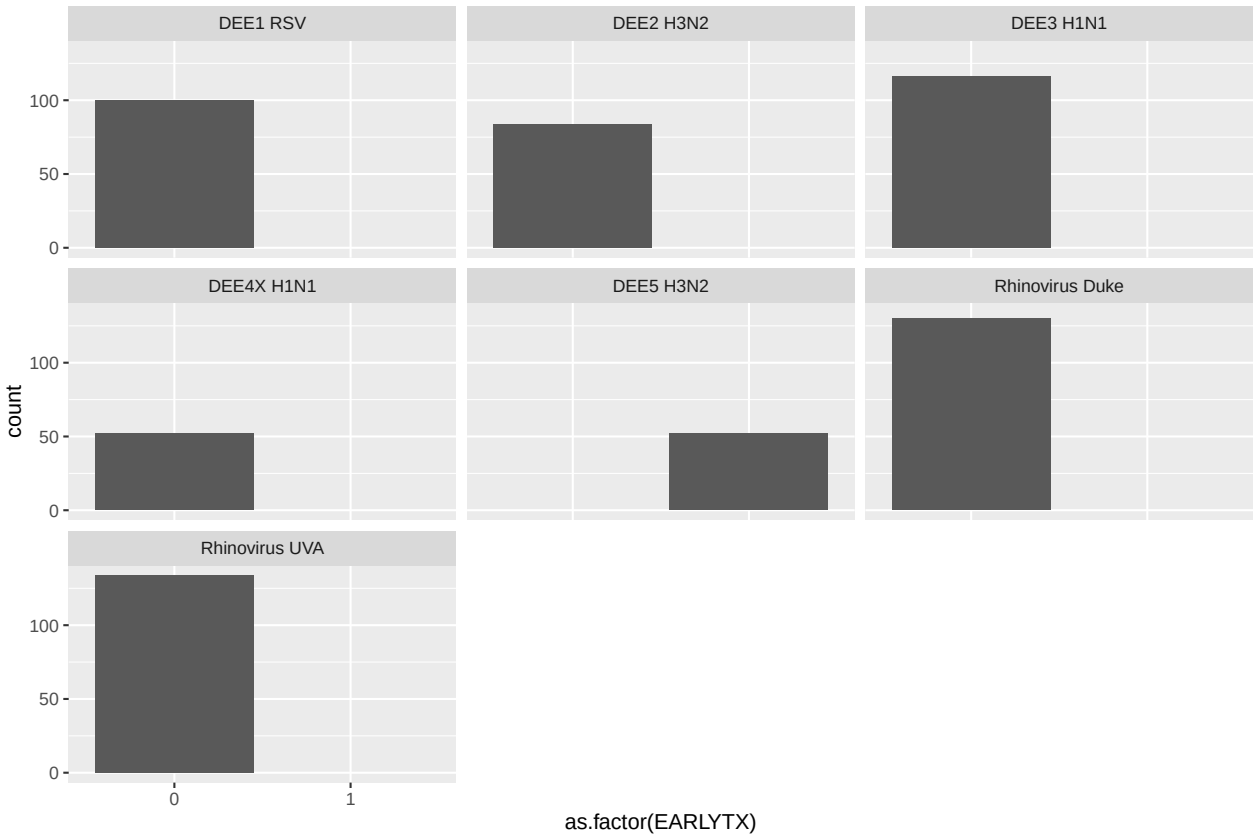
Clinical Features

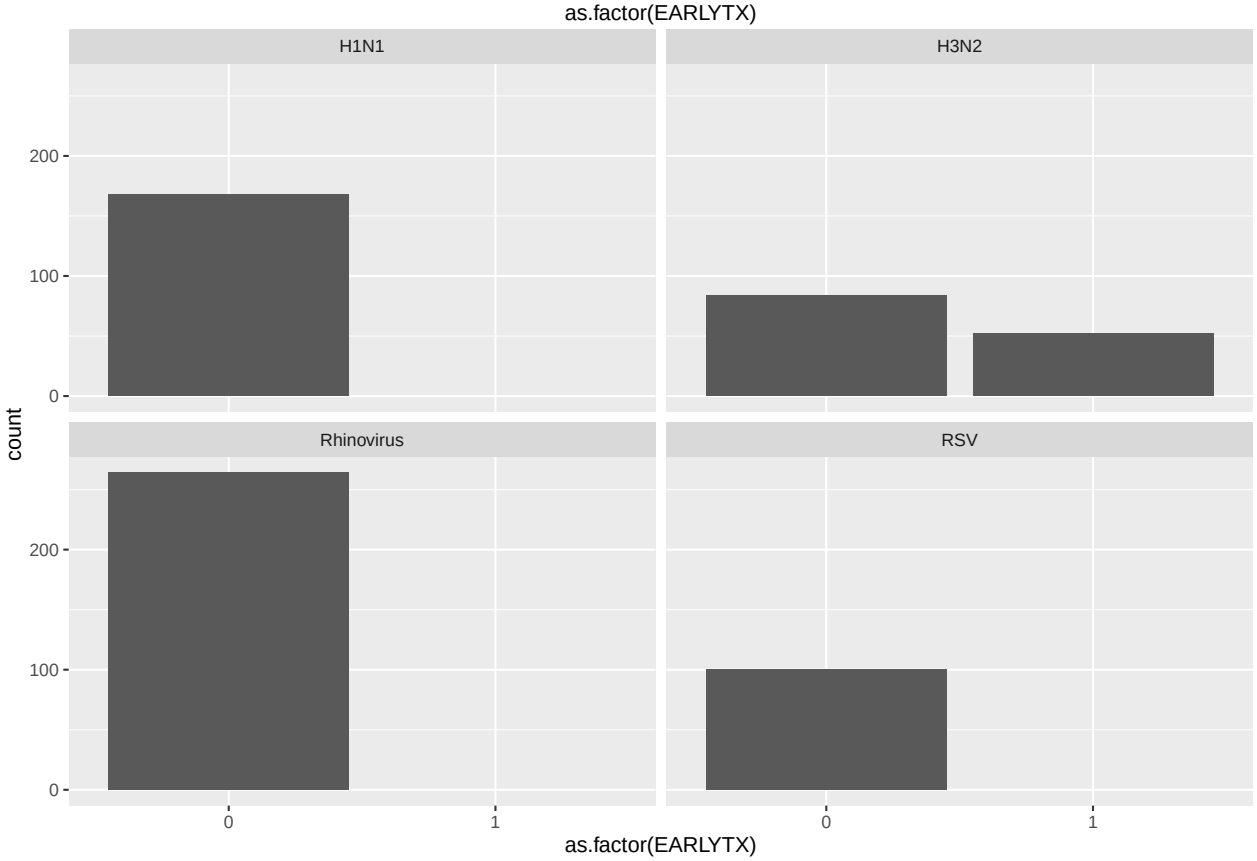
Timepoints are available in this file post-24h following viral challenge. Remove them prior to studying other clinical distributions.

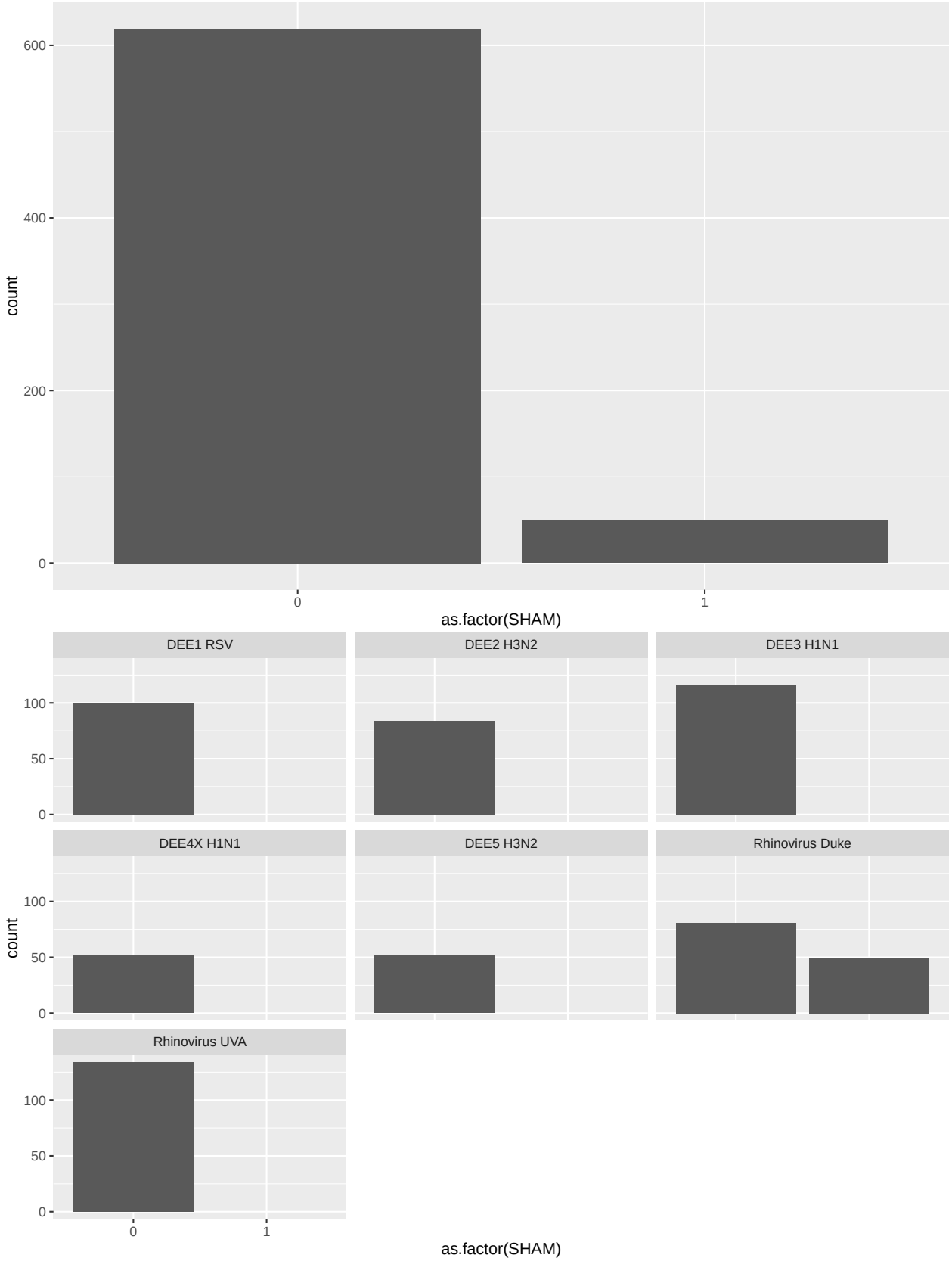

count
60
40
20
0
20
30
40
AGE
DEE1 RSV
DEE2 H3N2
DEE3 H1N1
DEE4X H1N1
DEE5 H3N2
Rhinovirus Duke
Rhinovirus UVA
count
40
30
20
10
0
20
30
40
AGE

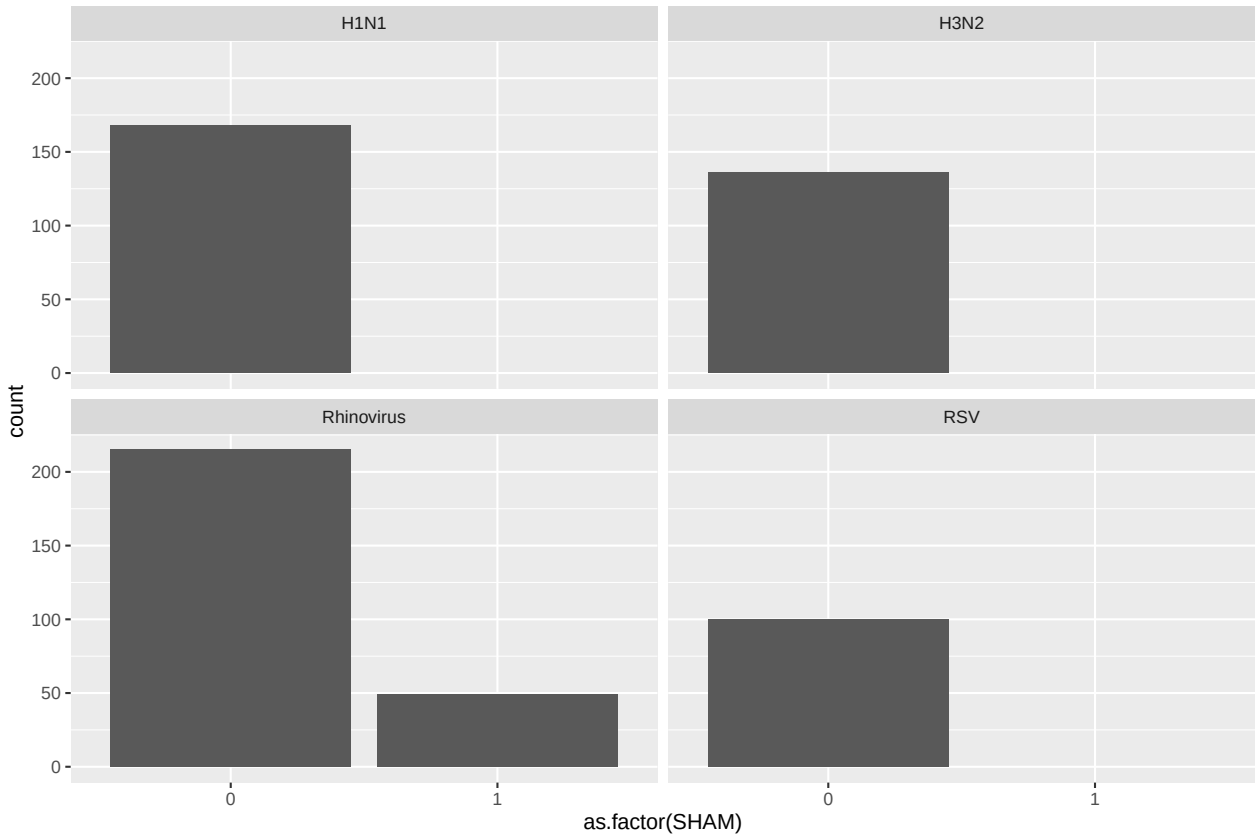

400
300
200
100
0
count
Female
Male
GENDER
DEE1 RSV
DEE2 H3N2
DEE3 H1N1
DEE4X H1N1
DEE5 H3N2
Rhinovirus Duke
Rhinovirus UVA
80
60
40
20
0
count
Female
Male
GENDER

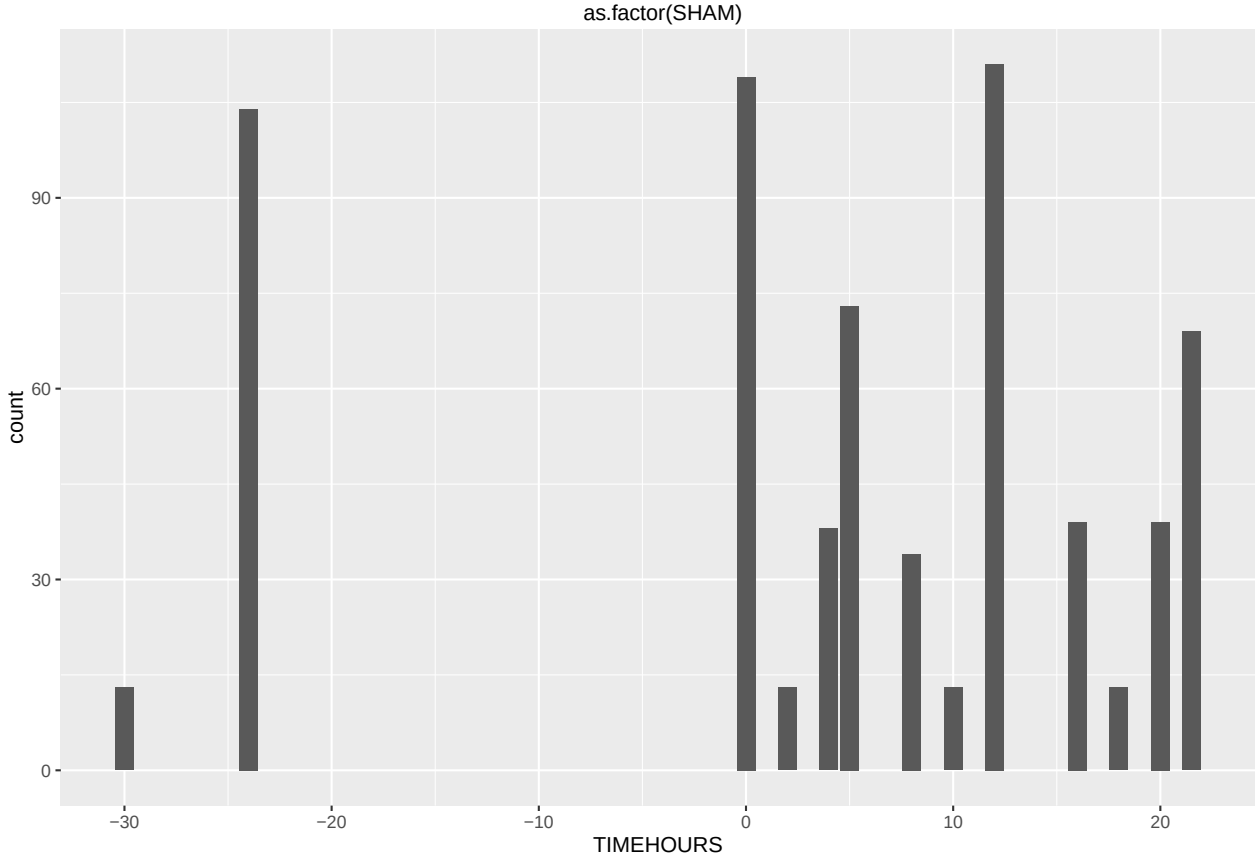

H1N1
H3N2
Rhinovirus
RSV
count
150
100
50
0
Female
Male
GENDER
600
400
200
0
1
as.factor(EARLYTX)

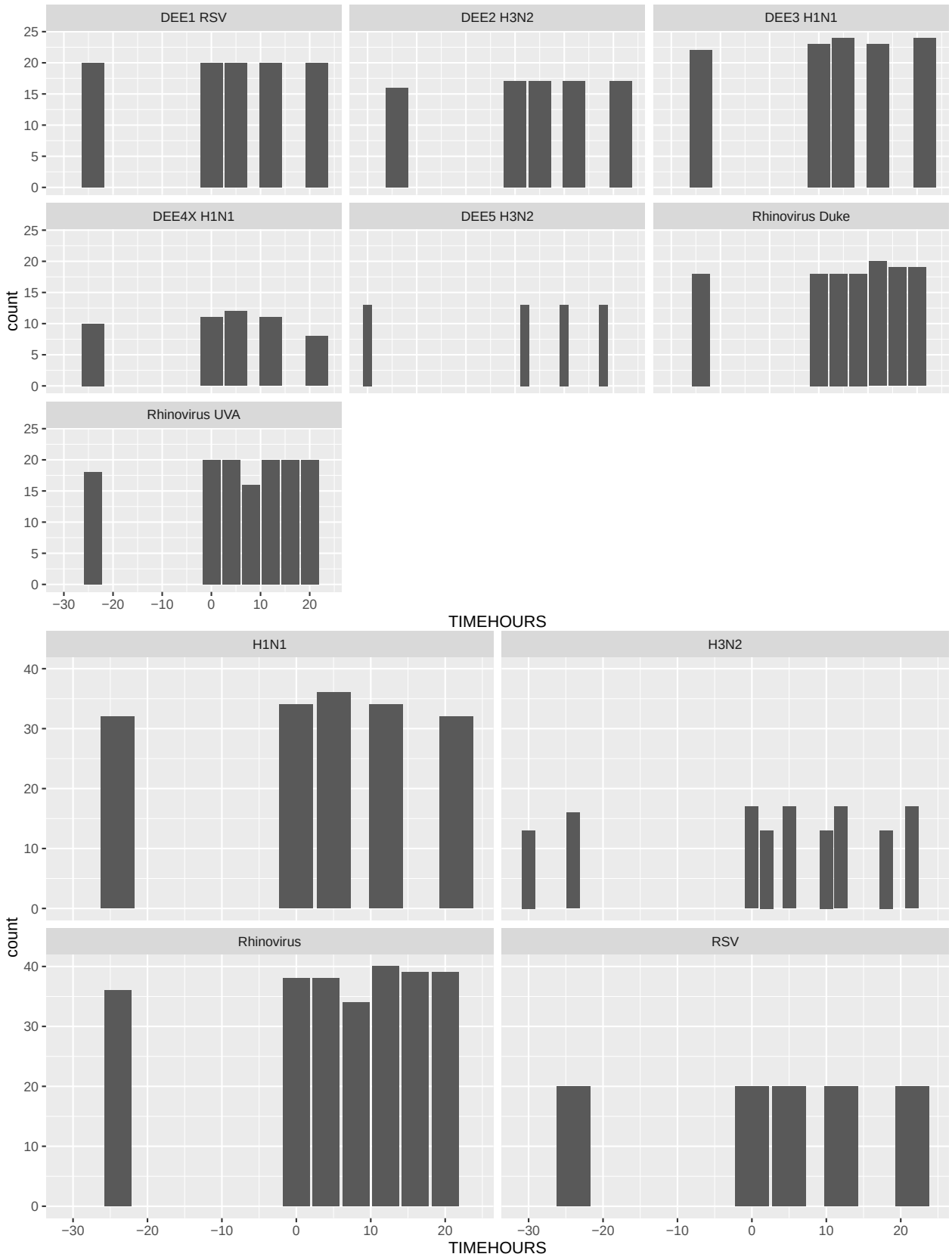

DEE1 RSV
DEE2 H3N2
DEE3 H1N1
DEE4X H1N1
DEE5 H3N2
Rhinovirus Duke
Rhinovirus UVA
count
100
50
0
0
1
as.factor(EARLYTX)


H1N1
H3N2
Rhinovirus
RSV
count
200
100
0
0
1
as.factor(EARLYTX)

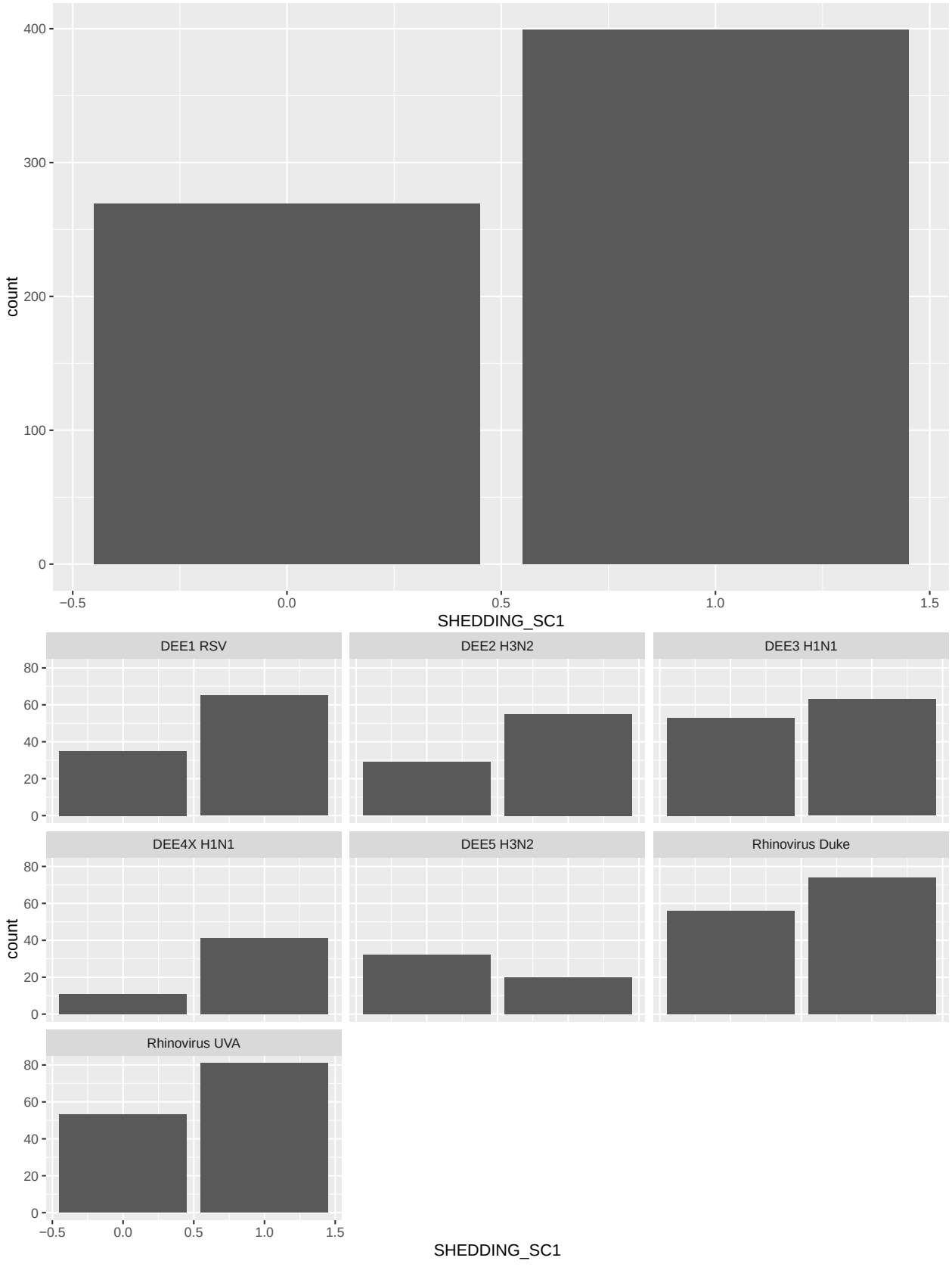

600
400
200
0
count
0
1
as.factor(SHAM)
DEE1 RSV
DEE2 H3N2
DEE3 H1N1
DEE4X H1N1
DEE5 H3N2
Rhinovirus Duke
Rhinovirus UVA
100
50
0
count
0
1
as.factor(SHAM)

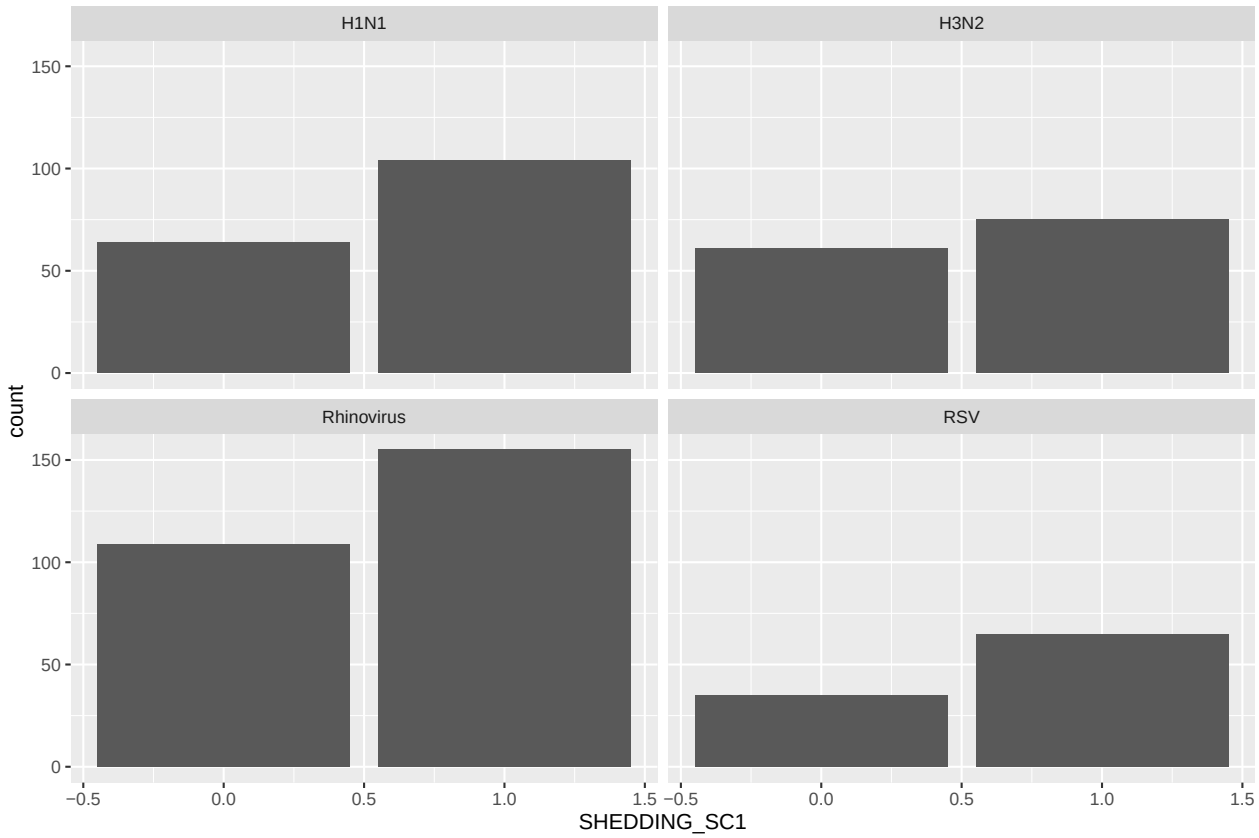

H1N1
H3N2
Rhinovirus
RSV
count
200
150
100
50
0
0
1
as.factor(SHAM)

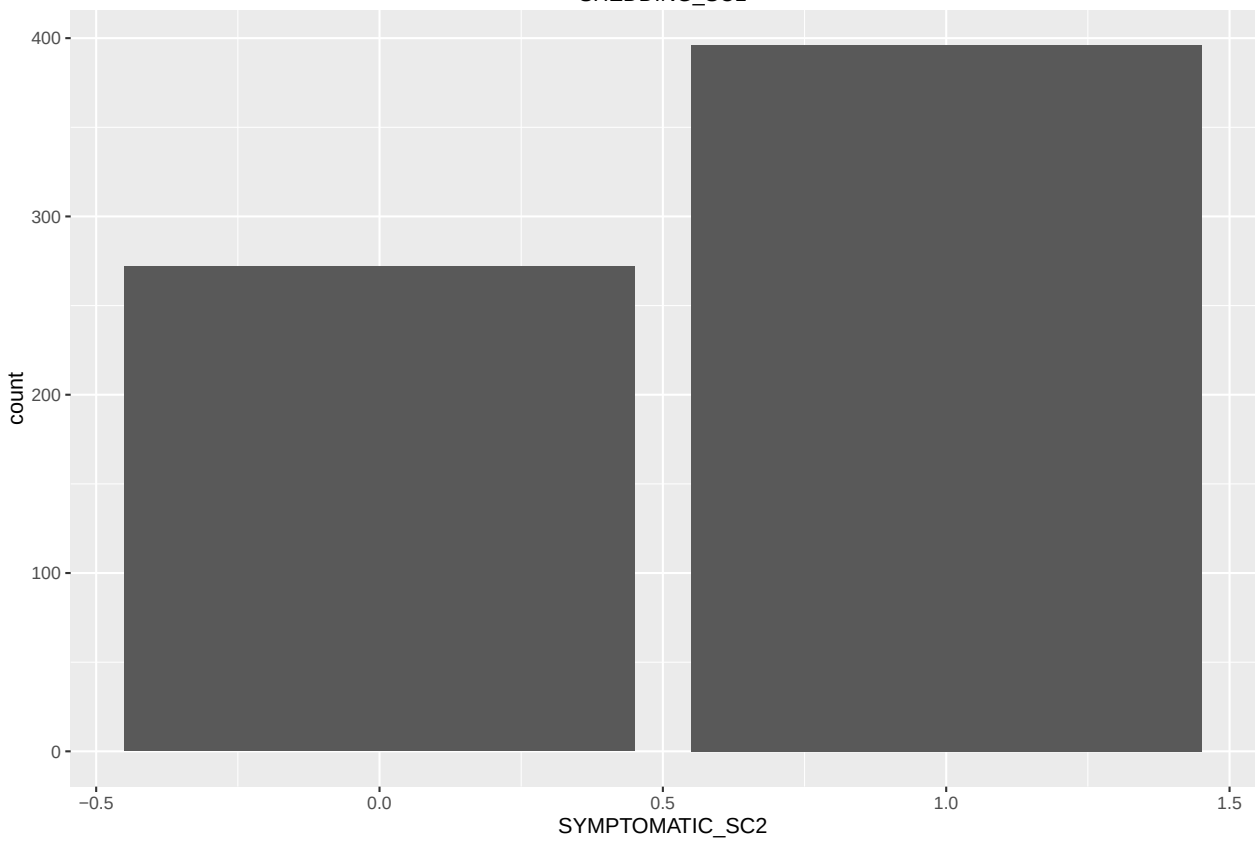

count
90
60
30
0
−30
−20
−10
0
10
20
TIMEHOURS

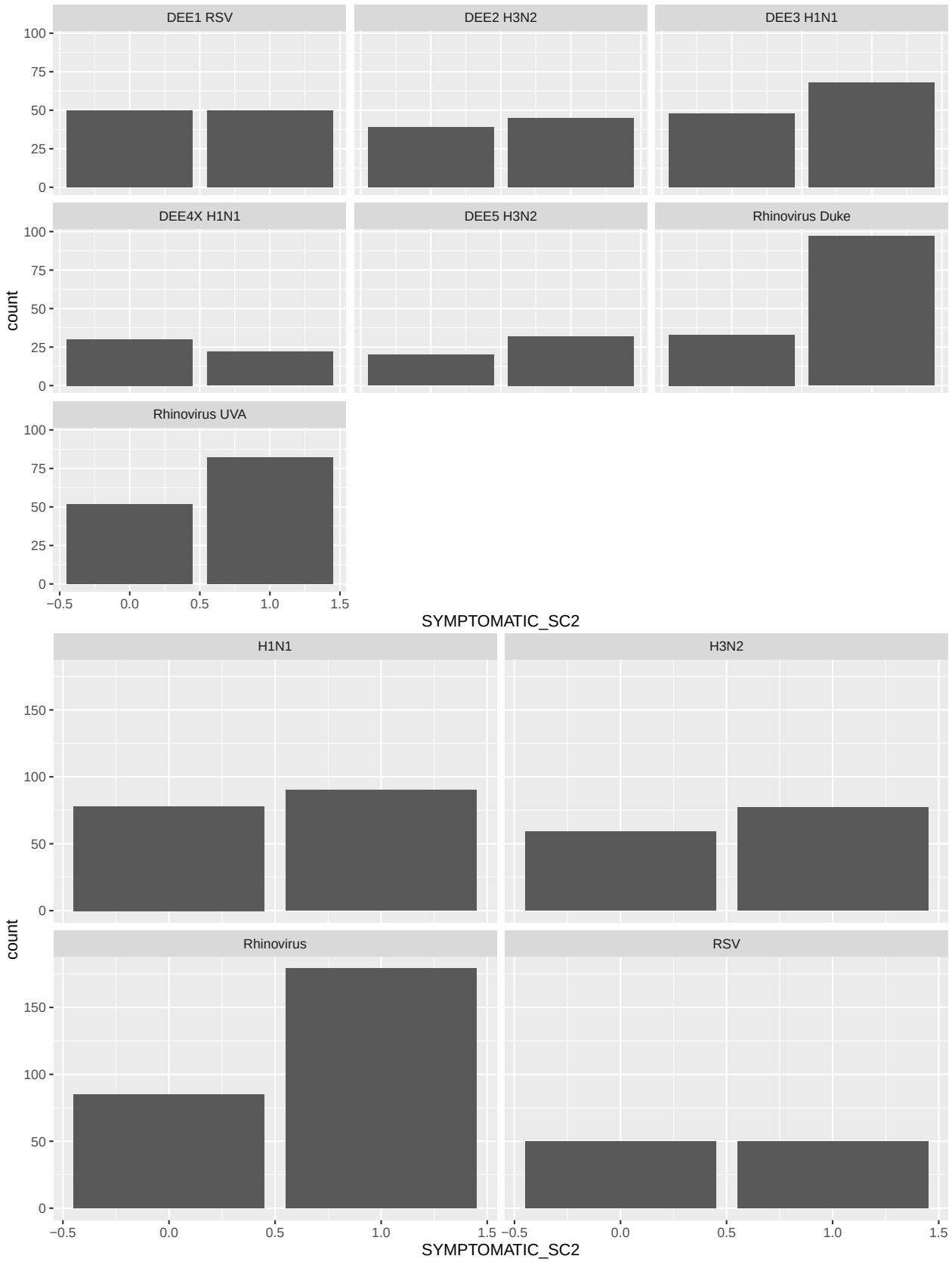

DEE1 RSV
DEE2 H3N2
DEE3 H1N1
DEE4X H1N1
DEE5 H3N2
Rhinovirus Duke
Rhinovirus UVA
count
TIMEHOURS
H1N1
H3N2
Rhinovirus
RSV
count
TIMEHOURS

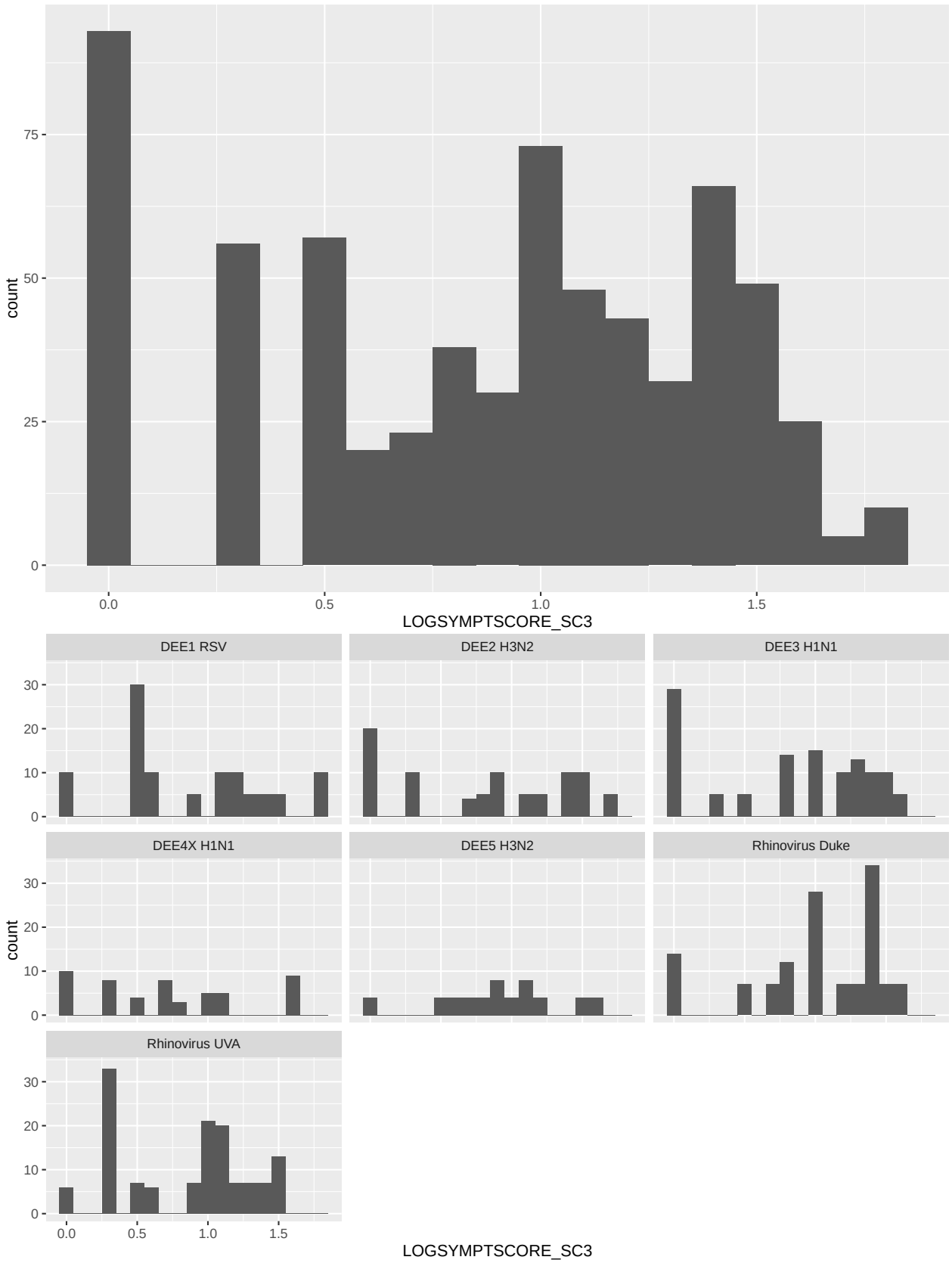
Clinical Outcomes

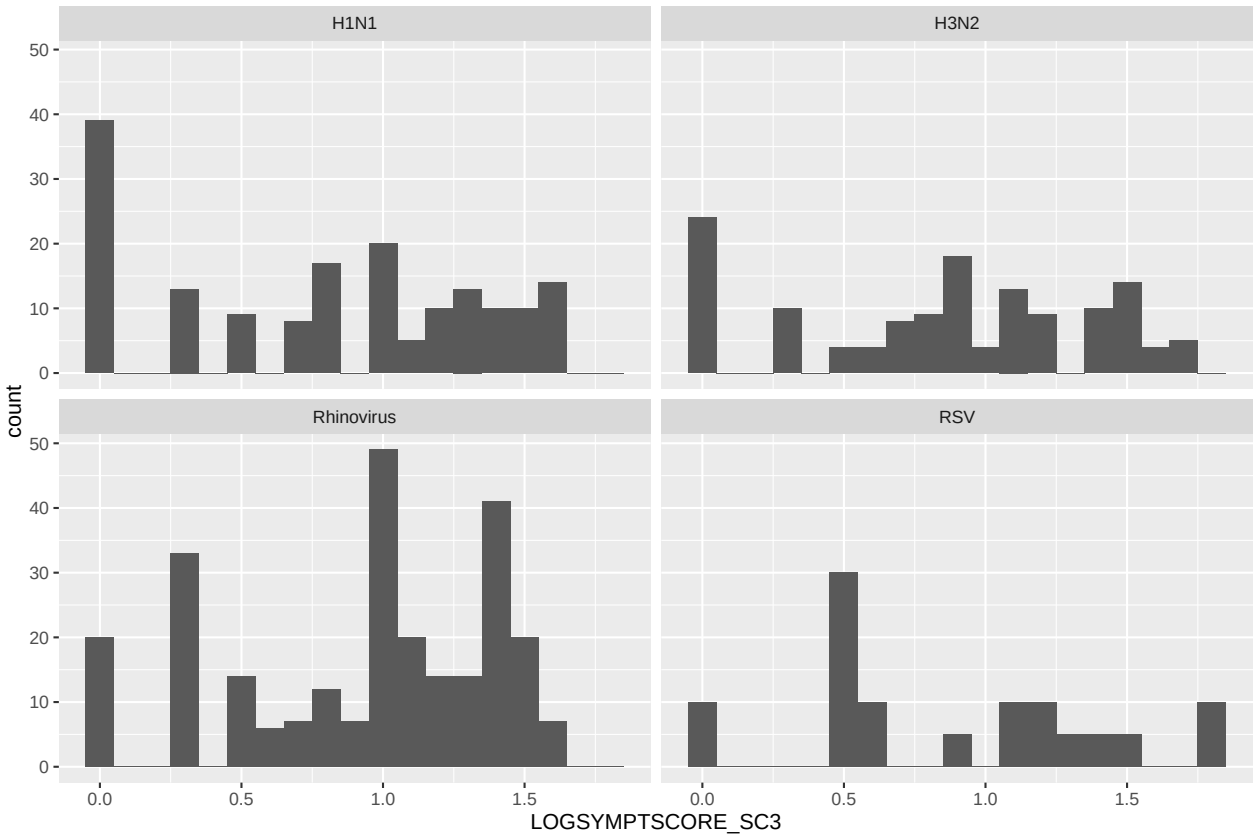



H1N1
H3N2
Rhinovirus
RSV
count
150
100
50
0
−0.5
0.0
0.5
1.0
1.5
SHEDDING_SC1

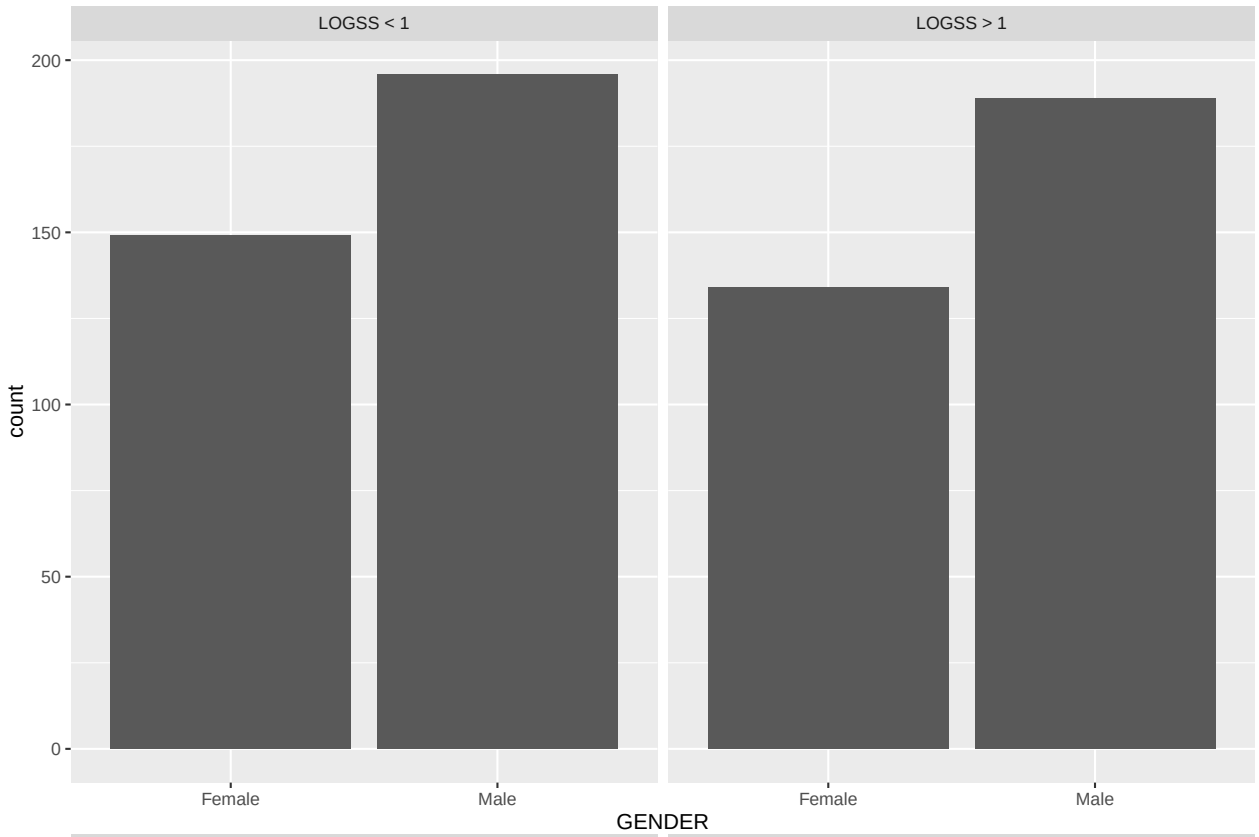

count
400
300
200
100
0
−0.5
0.0
0.5
1.0
1.5
SYMPTOMATIC_SC2

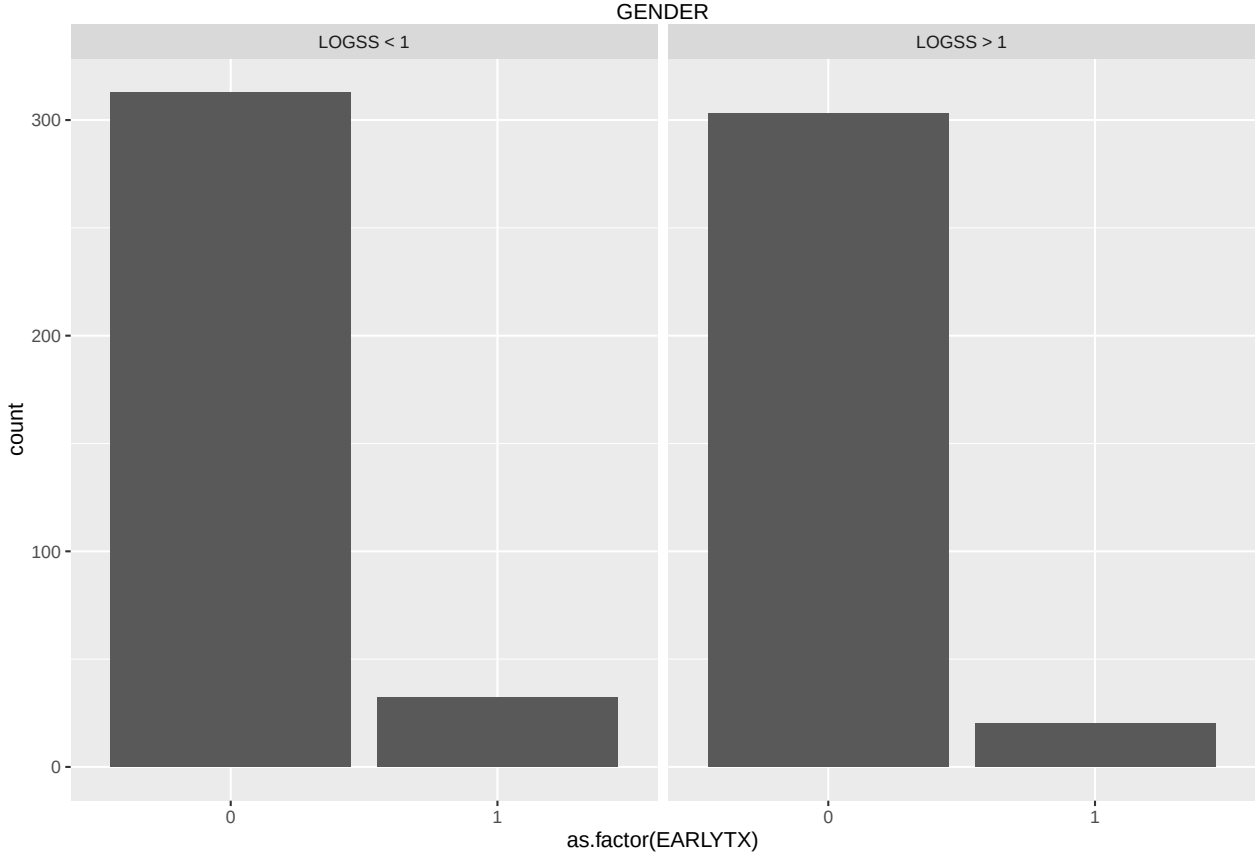

DEE1 RSV
DEE2 H3N2
DEE3 H1N1
DEE4X H1N1
DEE5 H3N2
Rhinovirus Duke
Rhinovirus UVA
count
SYMPTOMATIC_SC2
H1N1
H3N2
Rhinovirus
RSV
count
SYMPTOMATIC_SC2

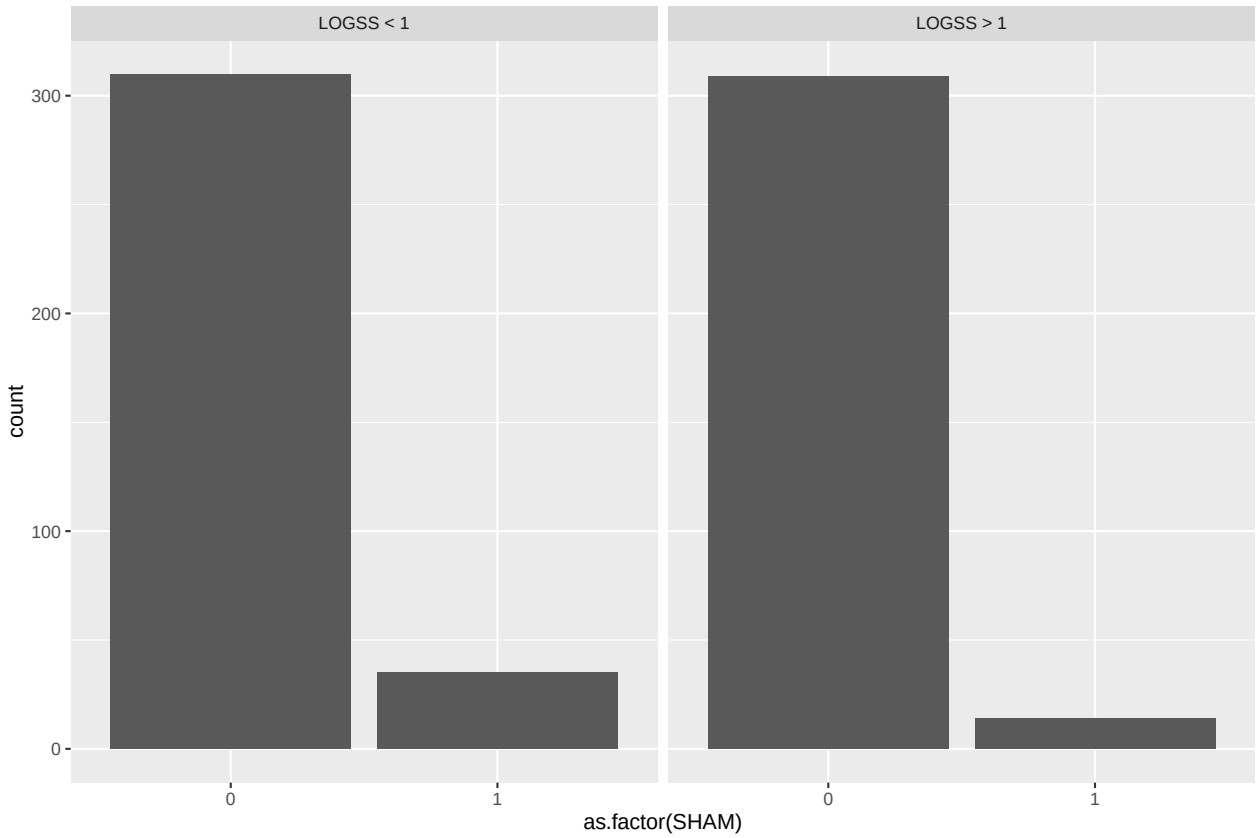

count
75
50
25
0
0.0
0.5
1.0
1.5
LOGSYMPTSCORE_SC3
DEE1 RSV
DEE2 H3N2
DEE3 H1N1
DEE4X H1N1
DEE5 H3N2
Rhinovirus Duke
Rhinovirus UVA
30
20
10
0
count
0.0
0.5
1.0
1.5
LOGSYMPTSCORE_SC3

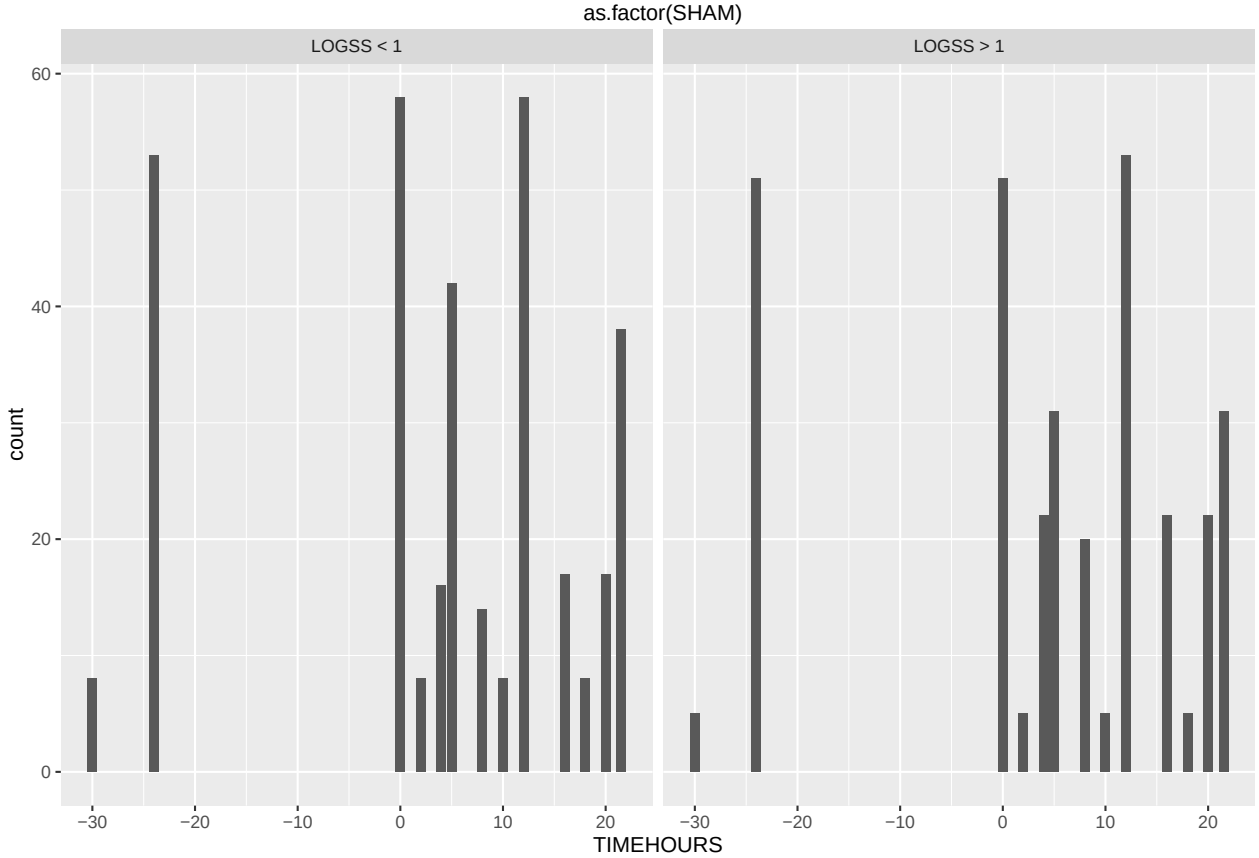


Features vs Outcomes


LOGSS < 1
LOGSS > 1
200
150
100
50
0
count
Female
Male
Female
Male
GENDER


LOGSS < 1
LOGSS > 1
300
200
100
0
count
0
1
0
1
as.factor(EARLYTX)


LOGSS < 1
LOGSS > 1
count
300
200
100
0
0
1
as.factor(SHAM)


LOGSS < 1
LOGSS > 1
count
60
40
20
0
−30
−20
−10
0
10
20
TIMEHOURS