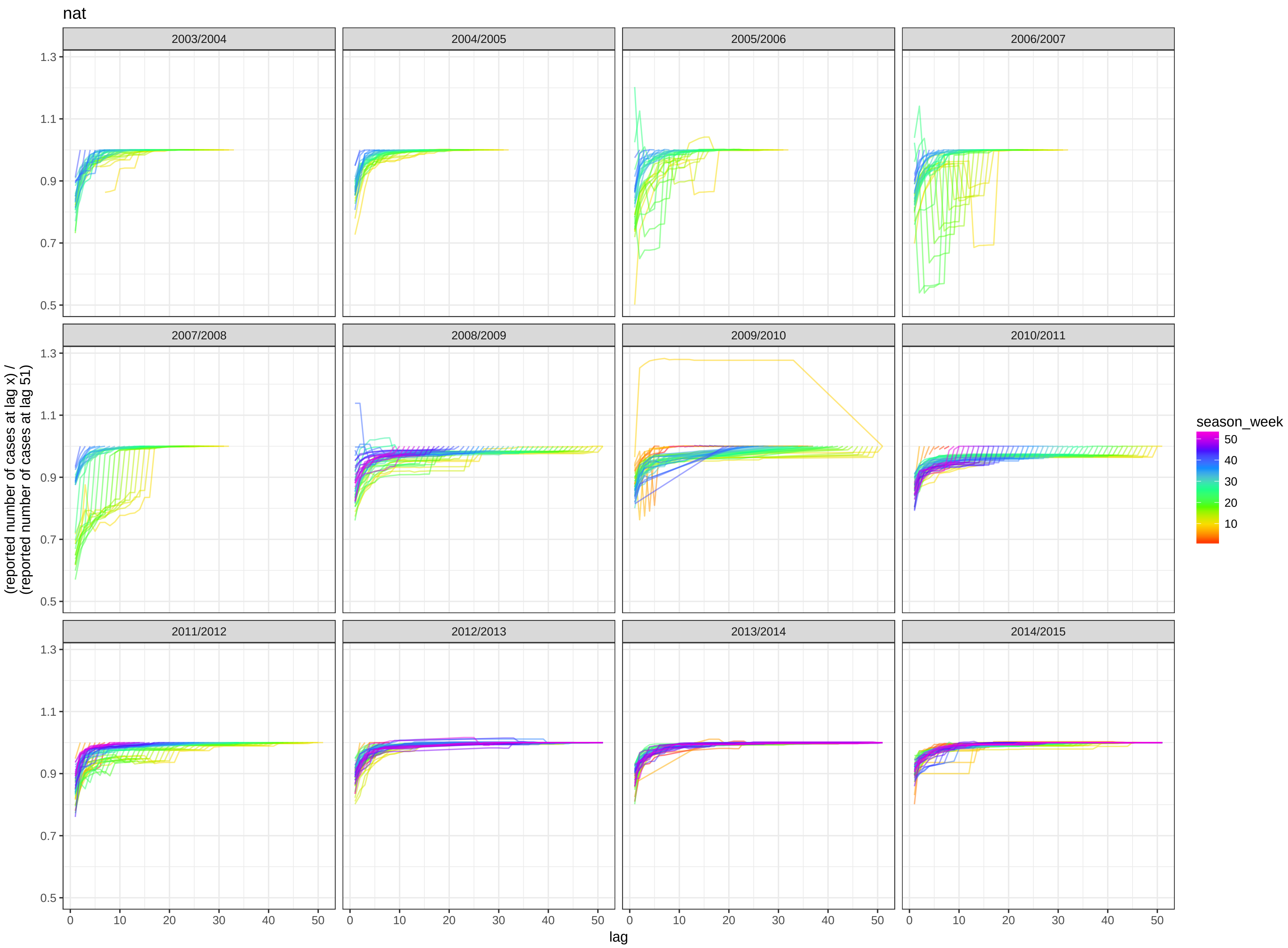
```
##
## Attaching package:  ’dplyr’
## The following objects are masked from ’package:plyr’:
##
##    arrange, count, desc, failwith, id, mutate, rename, summarise,
##    summarize
## The following objects are masked from ’package:stats’:
##
##    filter, lag
## The following objects are masked from ’package:base’:
##
##    intersect, setdiff, setequal, union
##
## Attaching package:  ’lubridate’
## The following object is masked from ’package:plyr’:
##
##    here
## The following object is masked from ’package:base’:
##
##    date
```

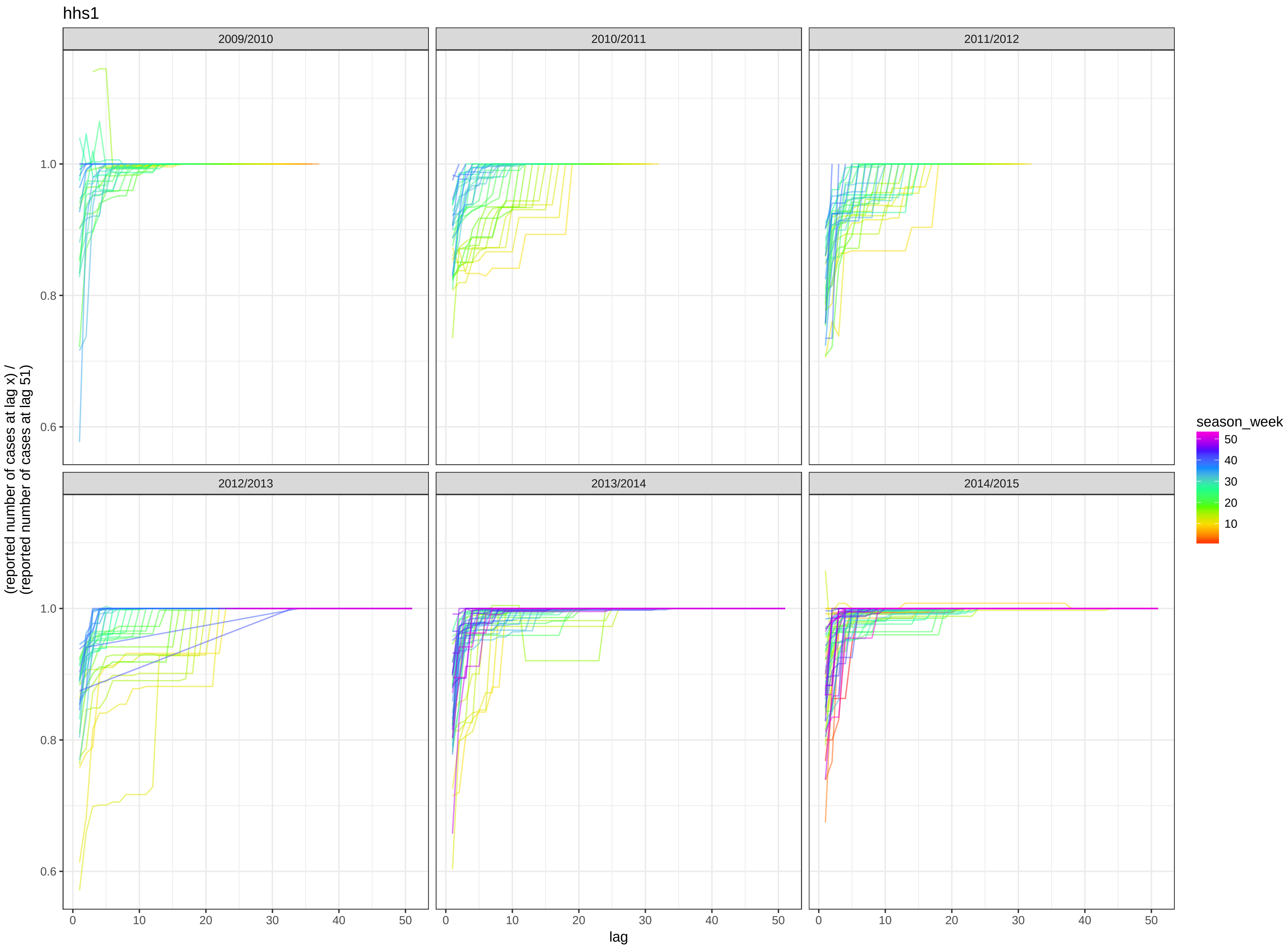
```
flu <- readRDS("data/flu_data_with_backfill.rds")
flu <- flu %>%
  mutate(
    year = substr(epiweek, 1, 4) %>% as.numeric(),
    week = substr(epiweek, 5, 6) %>% as.numeric(),
    report_year = substr(issue, 1, 4) %>% as.numeric(),
    report_week = substr(issue, 5, 6) %>% as.numeric(),
    epi_week_date = MMWRweek::MMWRweek2Date(
      MMWRyear = year,
      MMWRweek = week
    ),
    report_week_date = MMWRweek::MMWRweek2Date(
      MMWRyear = report_year,
      MMWRweek = report_week
    )
  )

flu$season <- ifelse(
  flu$week <= 30,
  paste0(flu$year - 1, "/", flu$year),
  paste0(flu$year, "/", flu$year + 1)
)

## Season week column: week number within season
## weeks after week 30 get season_week = week - 30
## weeks before week 30 get season_week = week + (number of weeks in previous year) - 30
## This computation relies on the start_date function in package MMWRweek,
## which is not exported from that package's namespace!!!
flu$season_week <- ifelse(
  flu$week <= 30,
  flu$week + MMWRweek(MMWRweek:::start_date(flu$year) - 1)$MMWRweek - 30,
  flu$week - 30
)

flu <- flu %>%
  group_by_("region", "epi_week_date") %>%
  mutate(
    final_num_ili = num_ili[lag == max(lag)],
    diff_log_curr_final_num_ili = log(num_ili) - log(final_num_ili)
  ) %>%
  ungroup()
```

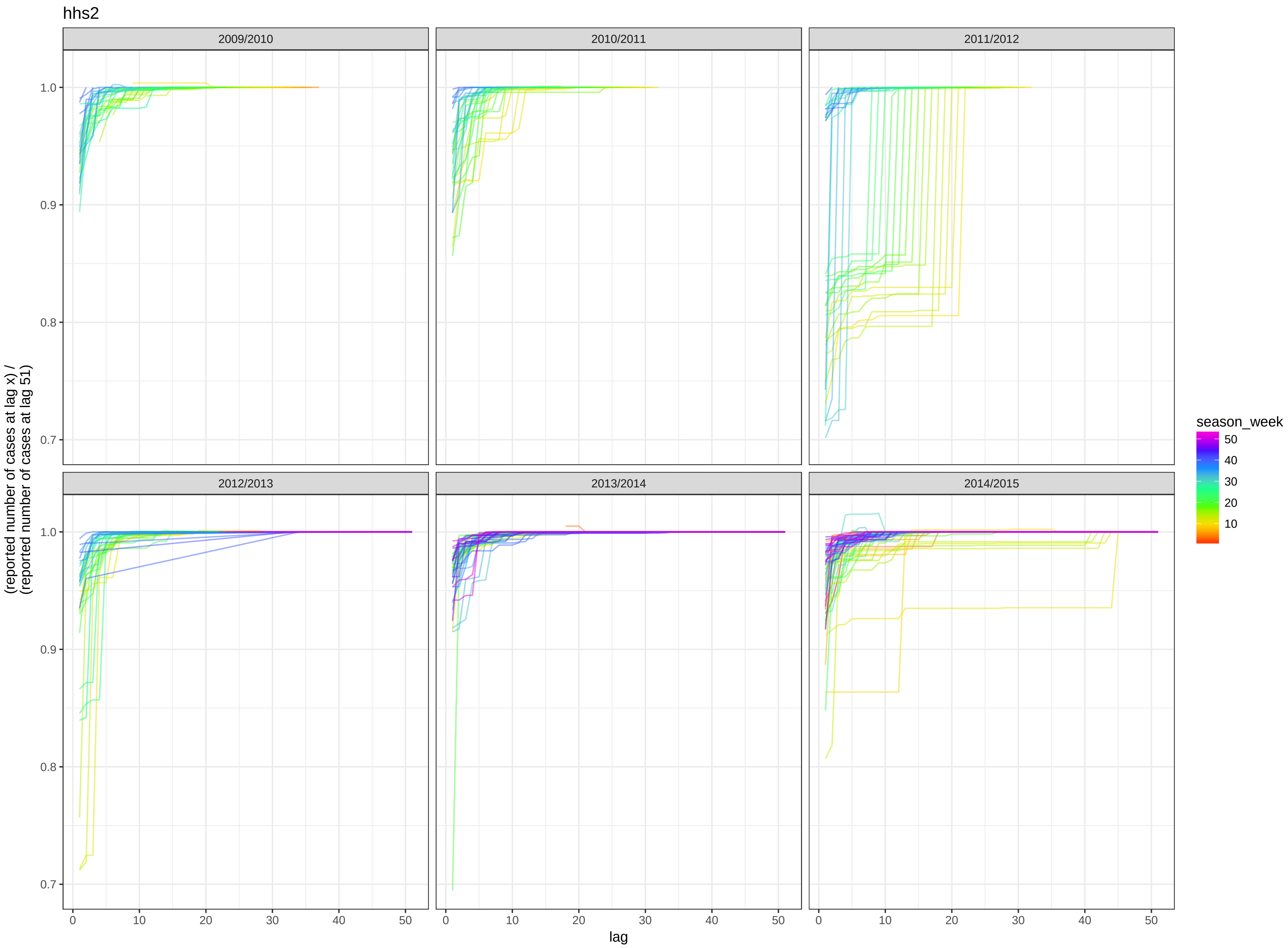
0.1 Visualizing backfill by region and season – ratios

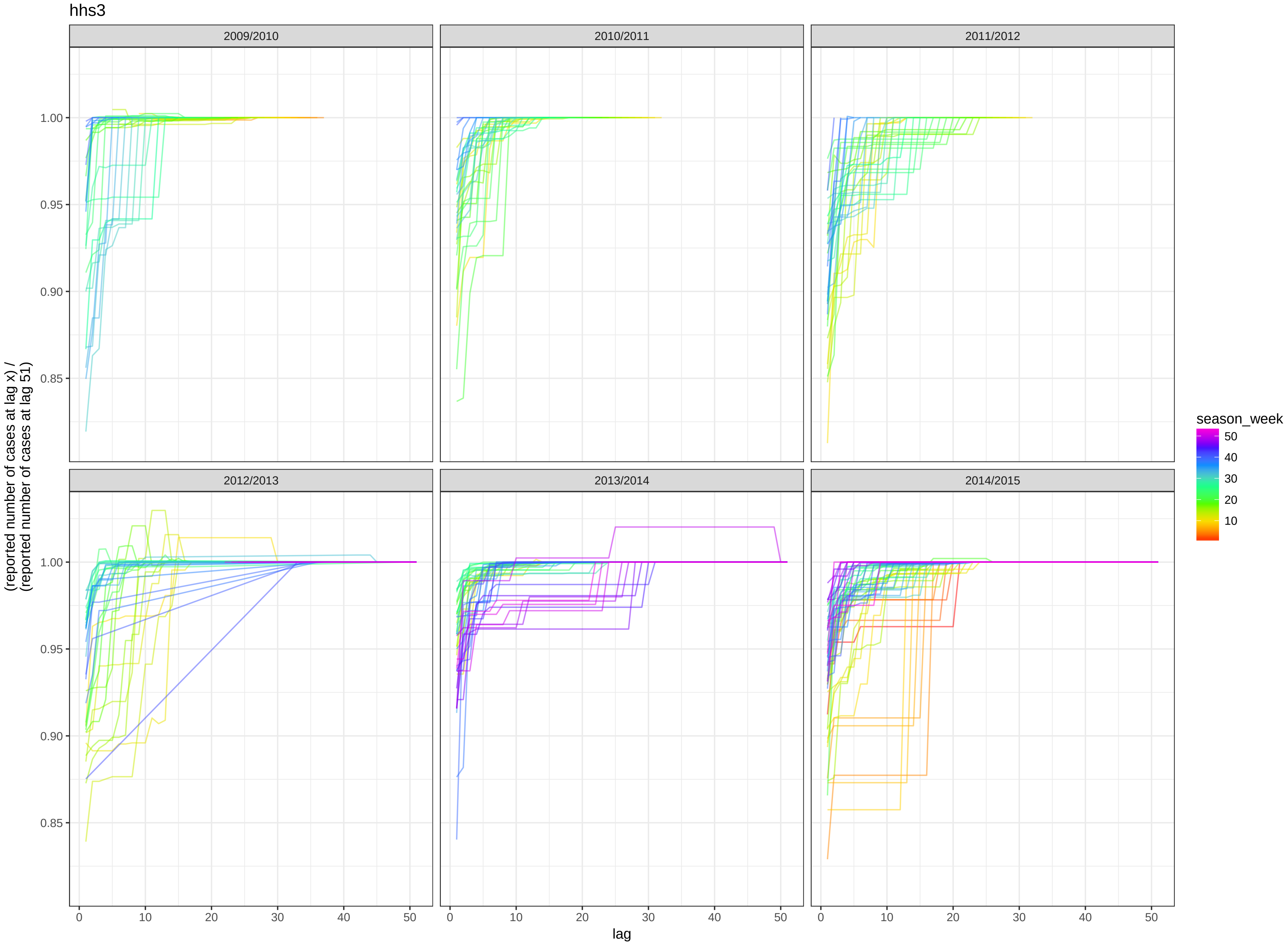
```
seasons_to_plot <- paste0(2003:2014, "/", 2004:2015)

for(region_val in unique(flu$region)) {
  p <- ggplot(data = flu %>% filter(region == region_val & season %in% seasons_to_plot)) +
    geom_line(
      aes(x = lag,
        y = exp(diff_log_curr_final_num_ili),
        group = epi_week_date,
        color = season_week),
      alpha = 0.5) +
    scale_color_gradientn(colors = rainbow(7)) +
    facet_wrap(~ season) +
    ylab("(reported number of cases at lag x) /\n(reported number of cases at lag 51)") +
    ggtitle(region_val) +
    theme_bw()
  print(p)
}
```

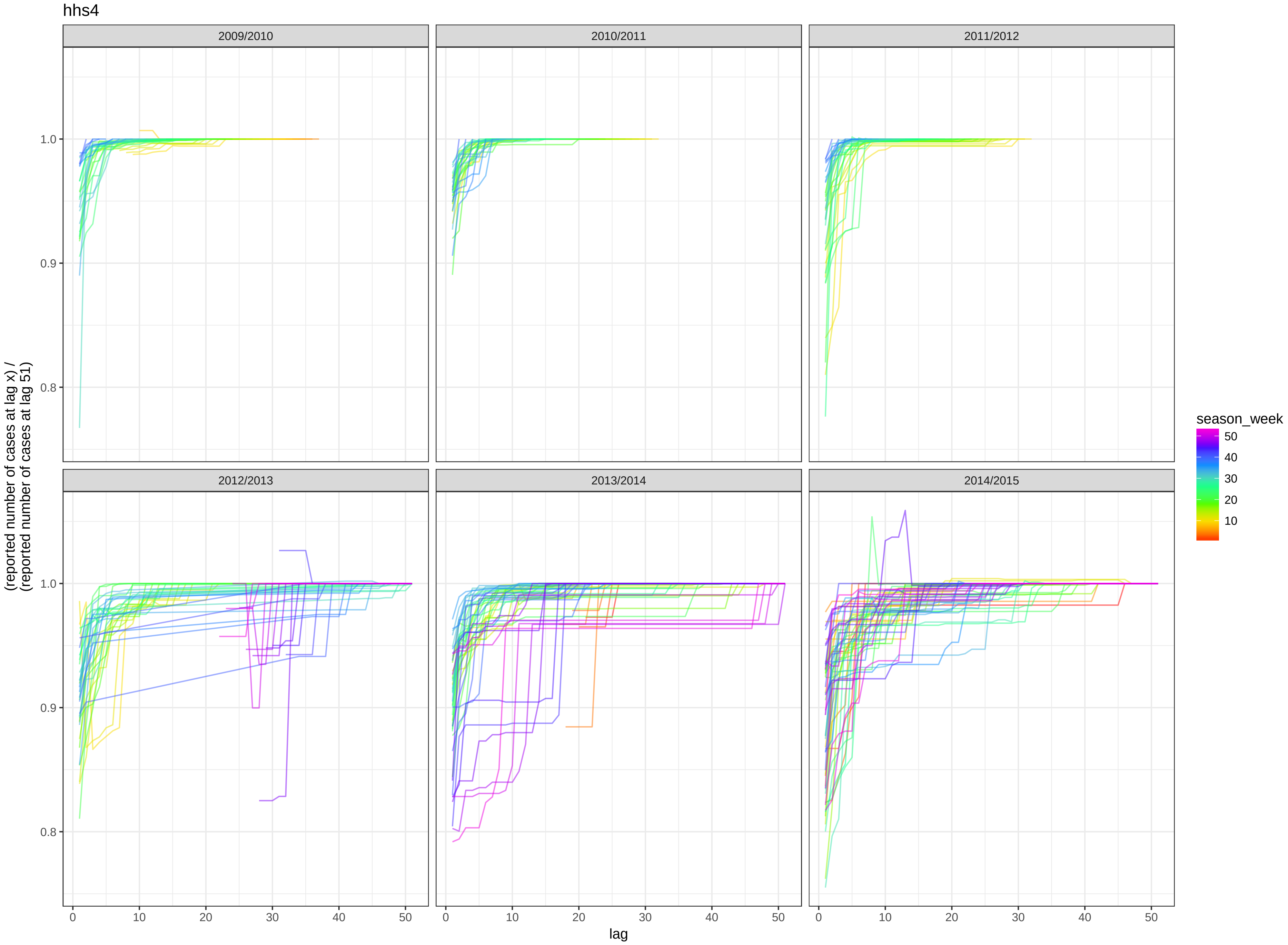


nat
2003/2004
2004/2005
2005/2006
2006/2007
2007/2008
2008/2009
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
1.3
1.1
0.9
0.7
0.5
0
10
20
30
40
50
lag
(reported number of cases at lag x) /
(reported number of cases at lag 51)
season_week
50
40
30
20
10

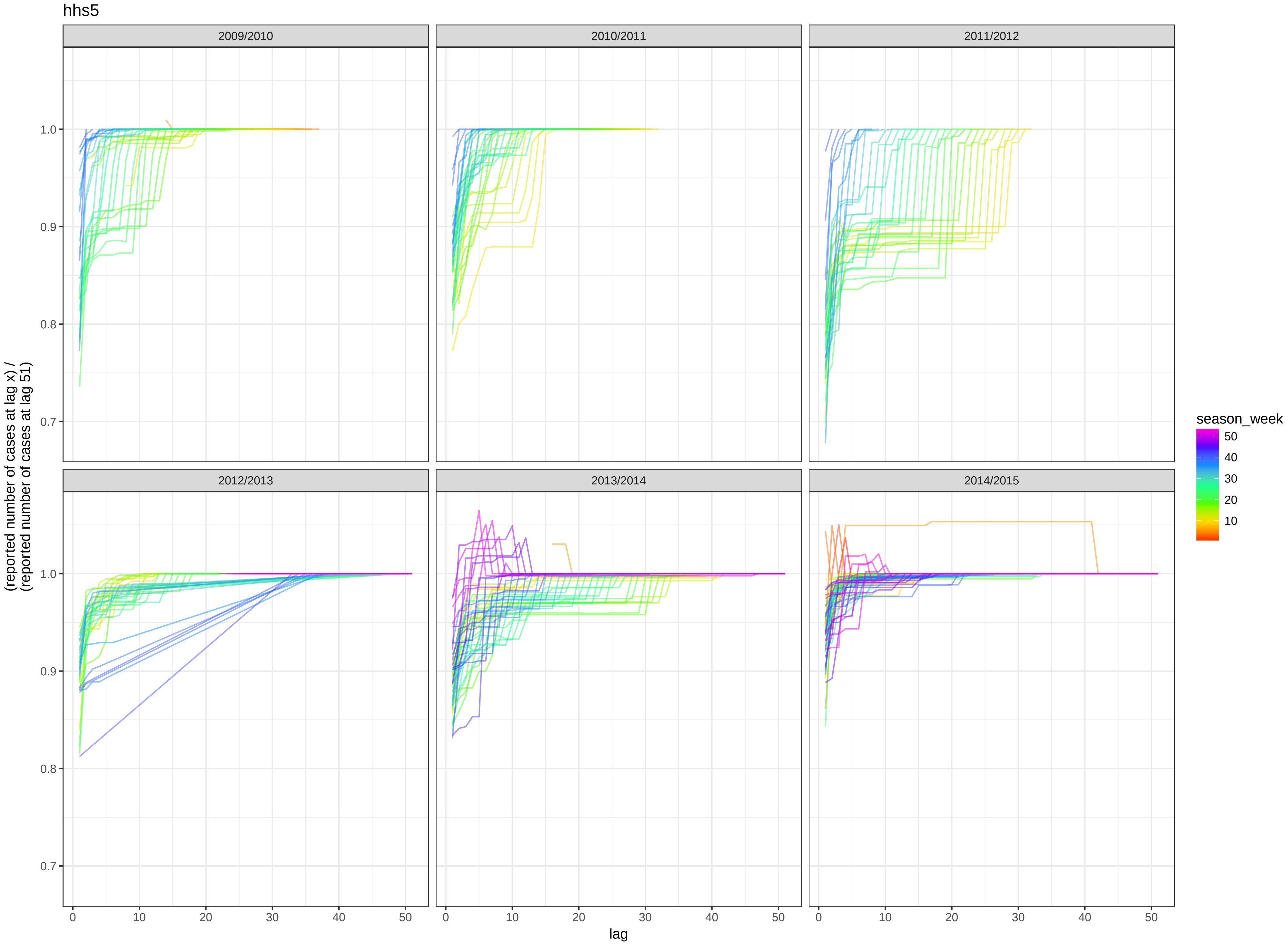

hhs1
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x) /
(reported number of cases at lag 51)
1.0
0.8
0.6
0
10
20
30
40
50
lag
season_week
50
40
30
20
10

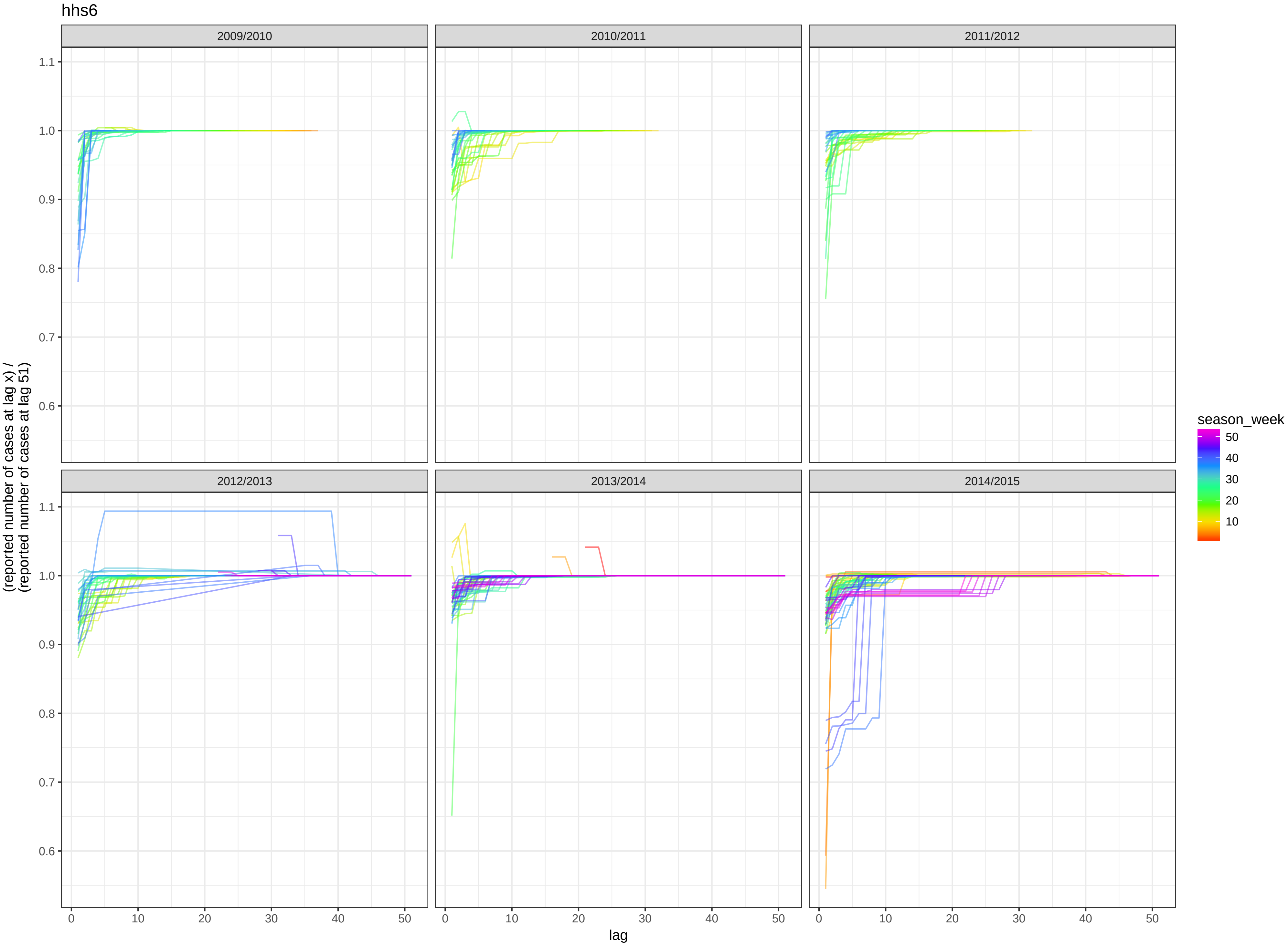

hhs2
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
1.0
0.9
0.8
0.7
(reported number of cases at lag x) /
(reported number of cases at lag 51)
0
10
20
30
40
50
lag
season_week
50
40
30
20
10

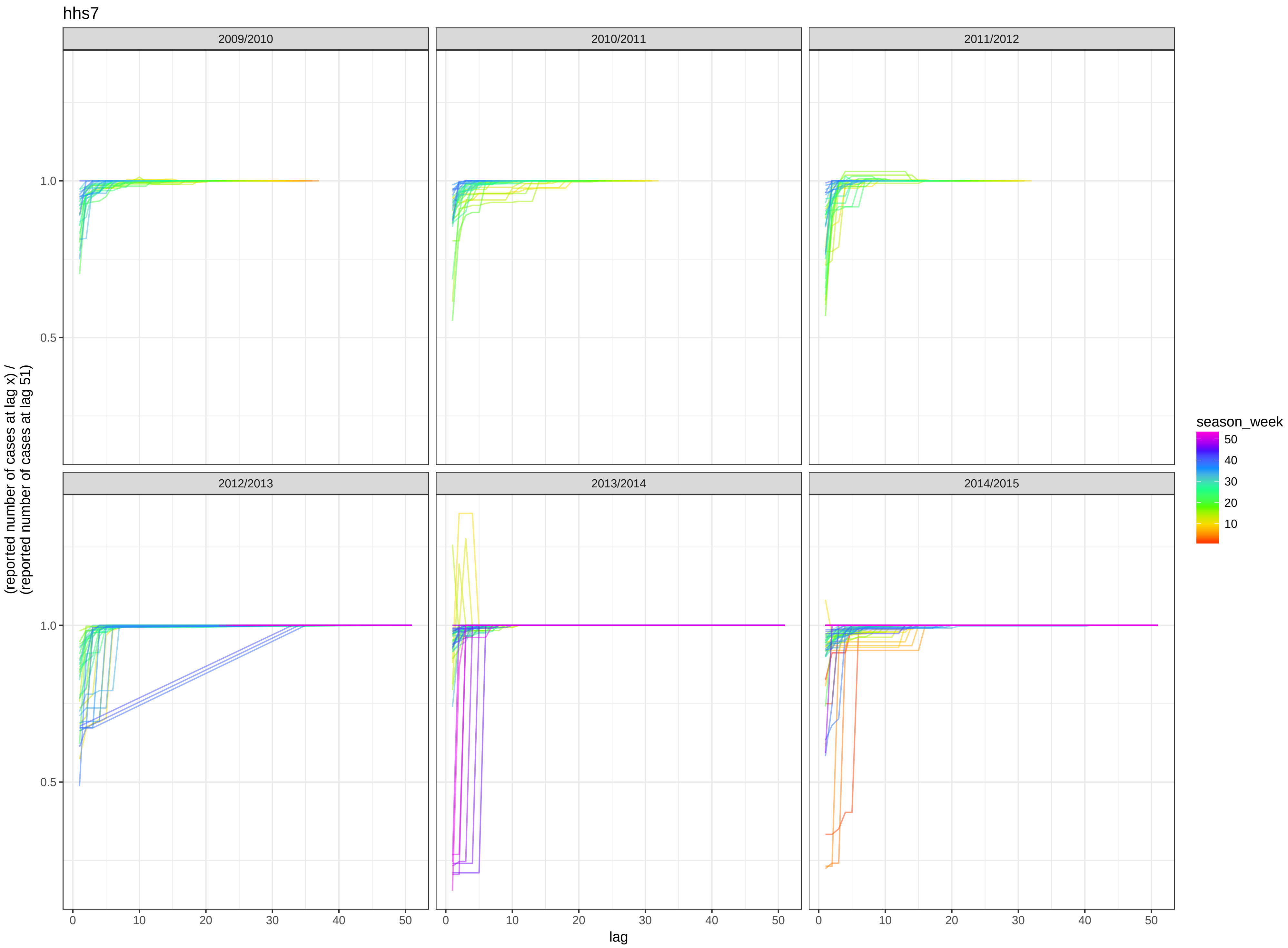

hhs3
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x) /
(reported number of cases at lag 51)
1.00
0.95
0.90
0.85
0
10
20
30
40
50
lag
season_week
50
40
30
20
10

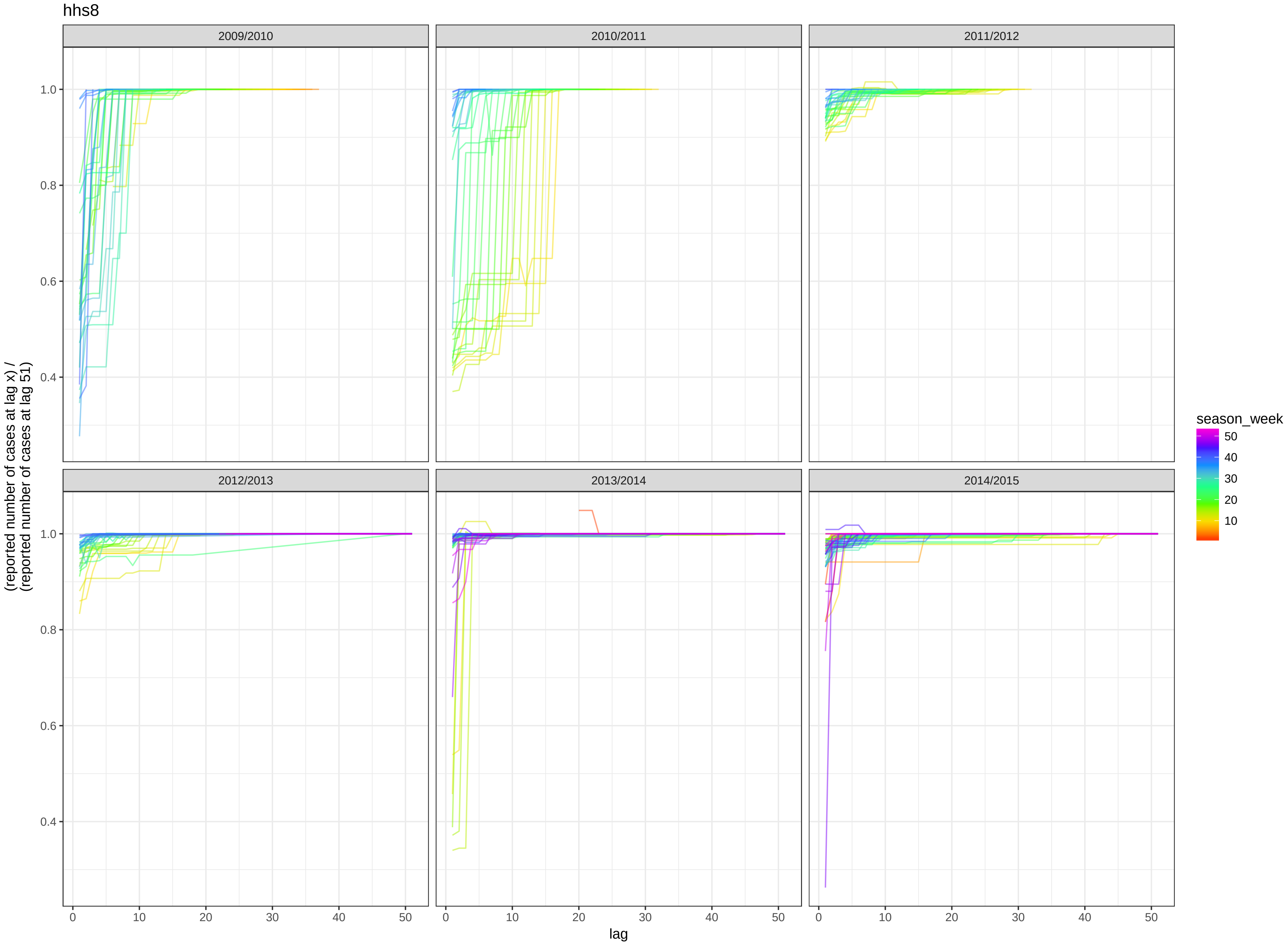


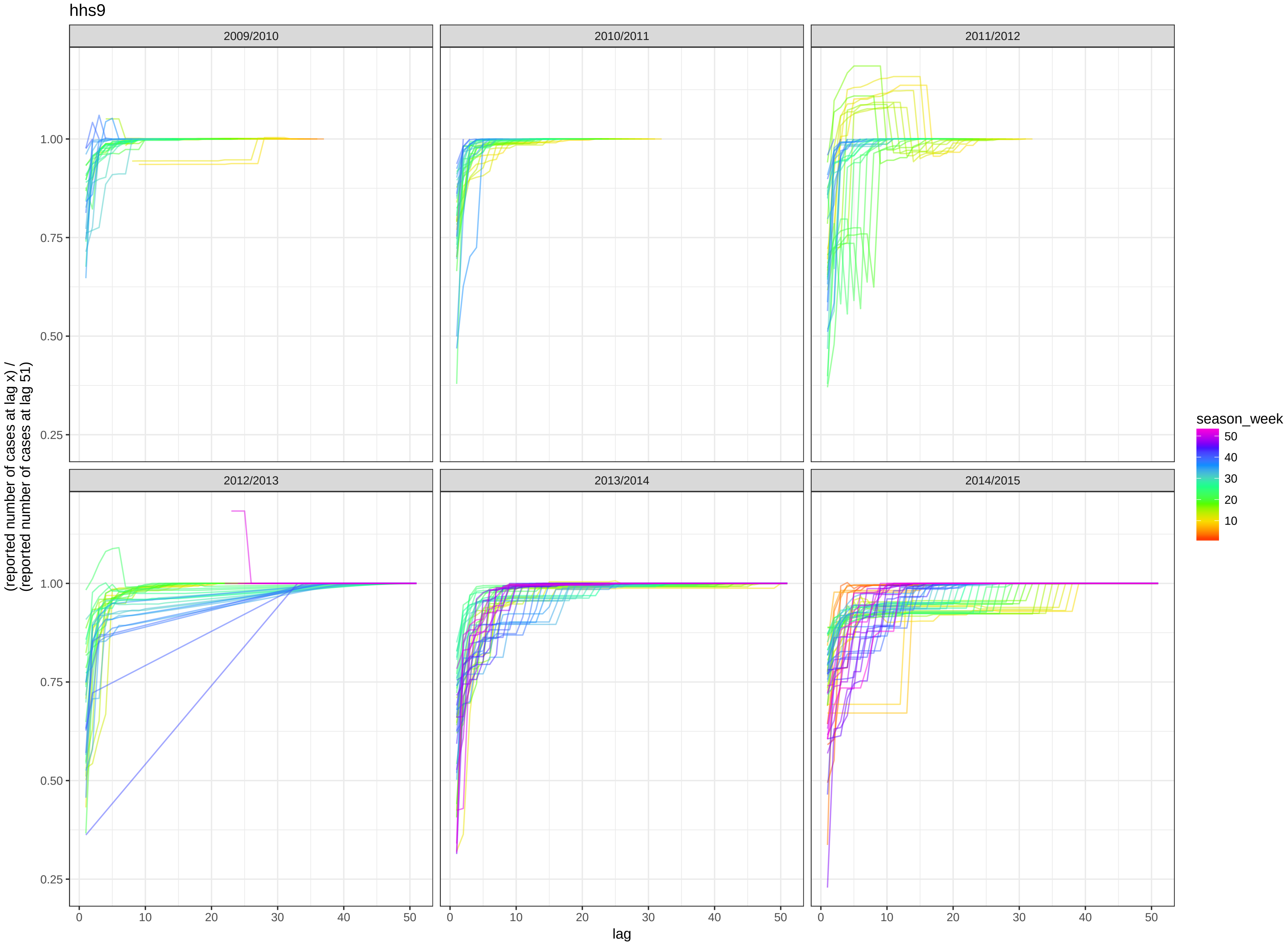
hhs4
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x) /
(reported number of cases at lag 51)
1.0
0.9
0.8
0
10
20
30
40
50
lag
season_week
50
40
30
20
10

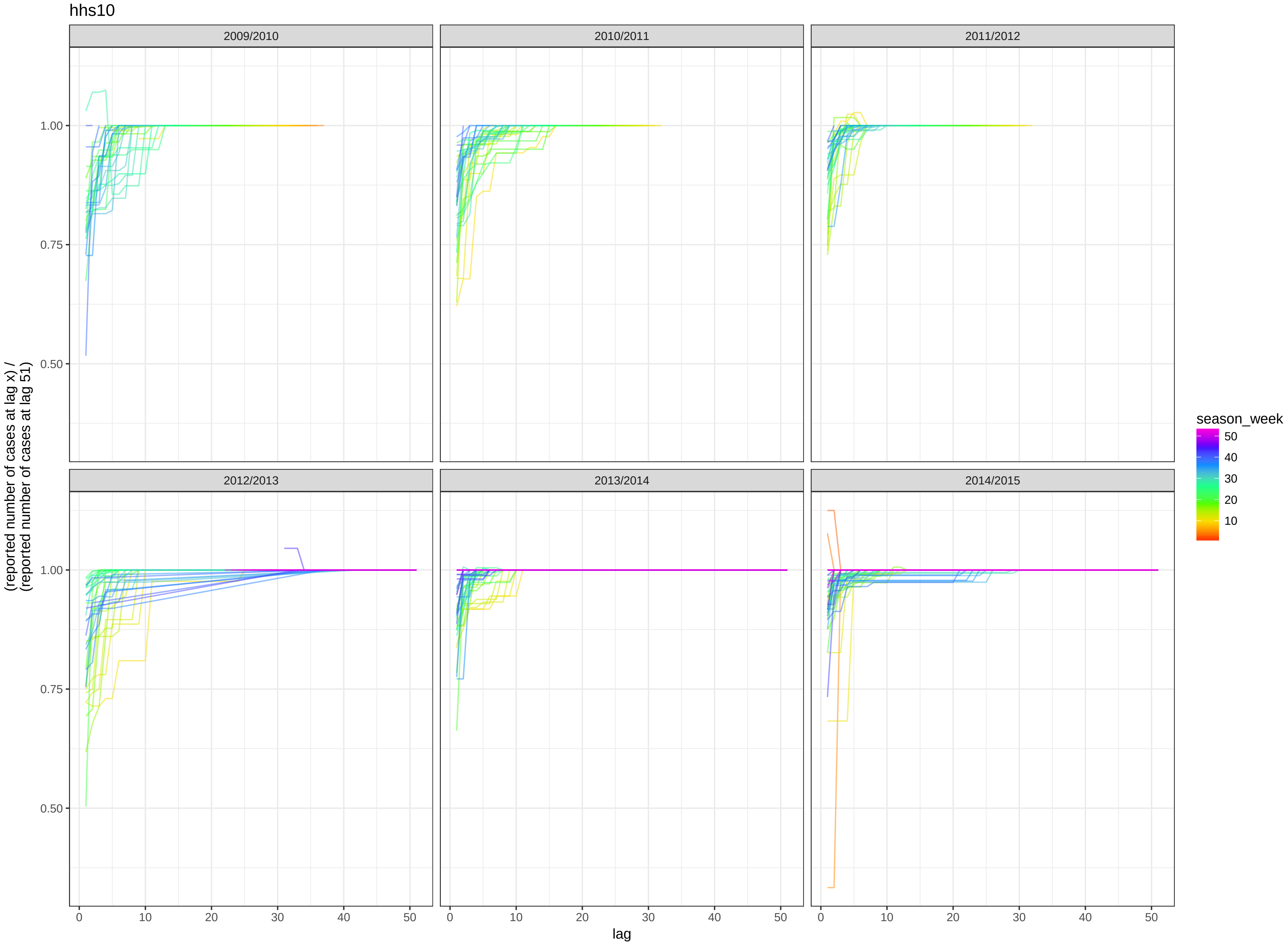

hhs5
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
1.0
0.9
0.8
0.7
0
10
20
30
40
50
lag
(reported number of cases at lag x) /
(reported number of cases at lag 51)
season_week
50
40
30
20
10


hhs6
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
1.1
1.0
0.9
0.8
0.7
0.6
(reported number of cases at lag x) /
(reported number of cases at lag 51)
0
10
20
30
40
50
lag
season_week
50
40
30
20
10


hhs7
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x) /
(reported number of cases at lag 51)
1.0
0.5
1.0
0.5
0
10
20
30
40
50
lag
season_week
50
40
30
20
10

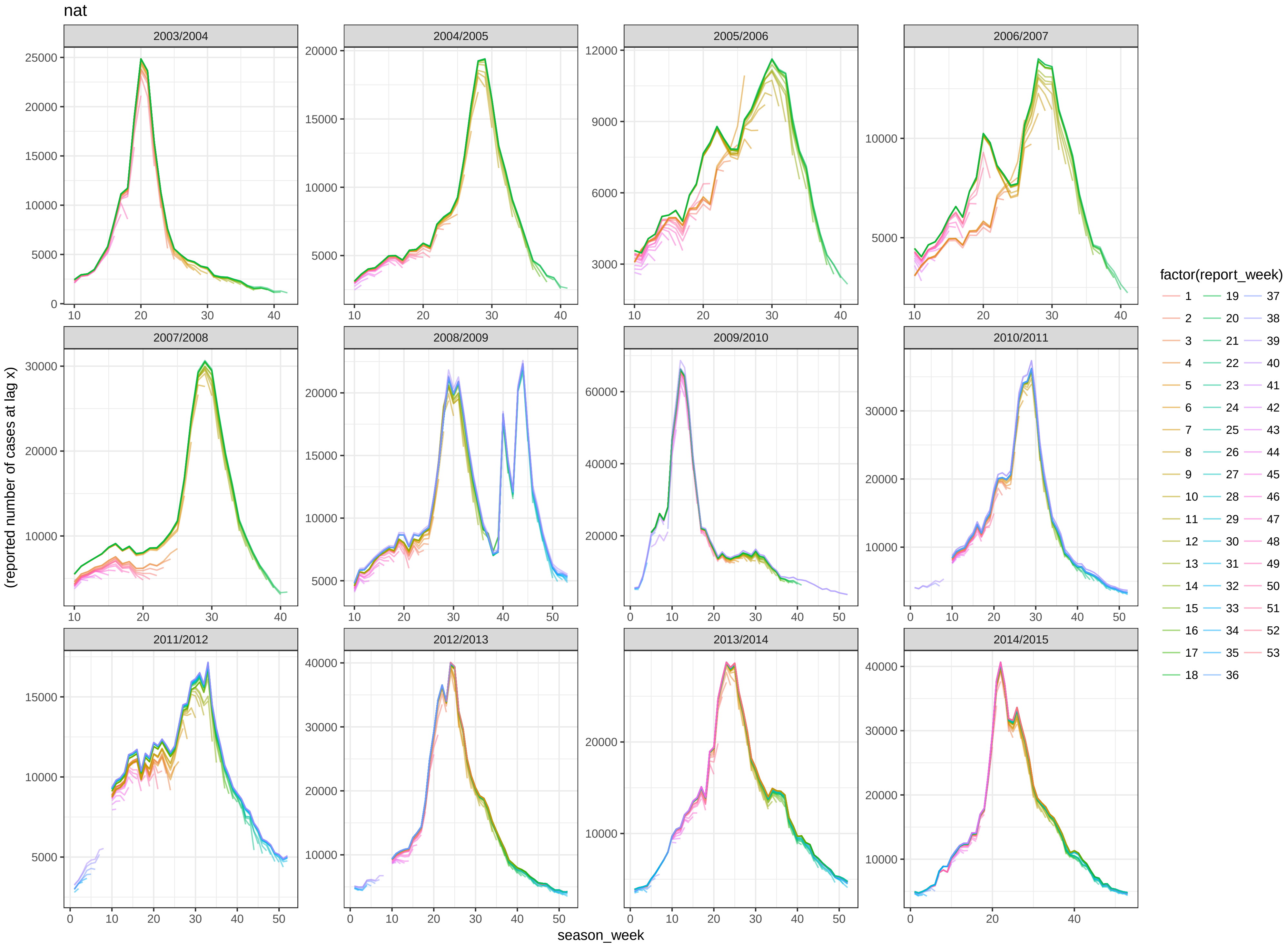

hhs8
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x) /
(reported number of cases at lag 51)
1.0
0.8
0.6
0.4
0
10
20
30
40
50
lag
season_week
50
40
30
20
10

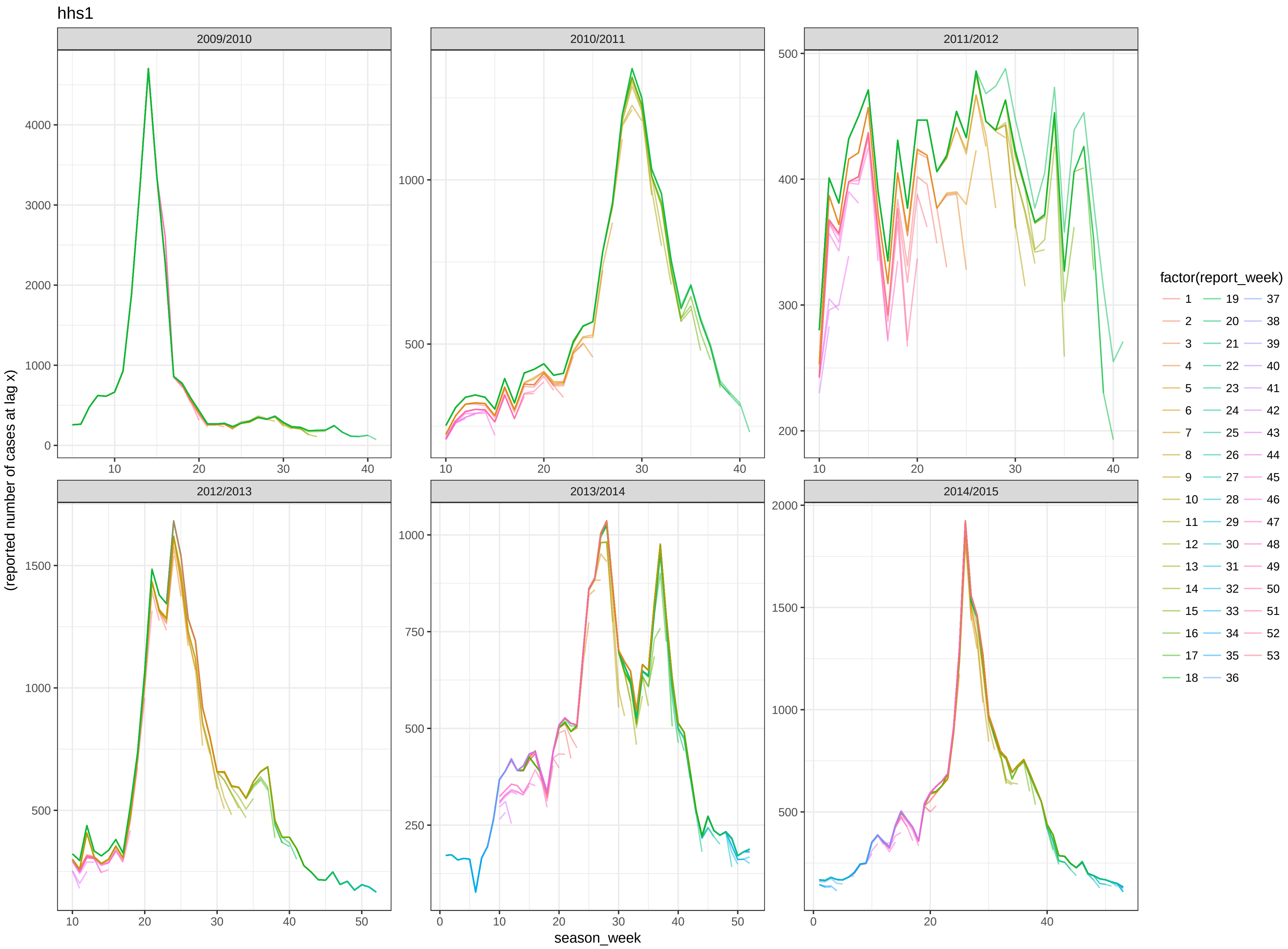

hhs9
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x) /
(reported number of cases at lag 51)
1.00
0.75
0.50
0.25
0
10
20
30
40
50
lag
season_week
50
40
30
20
10

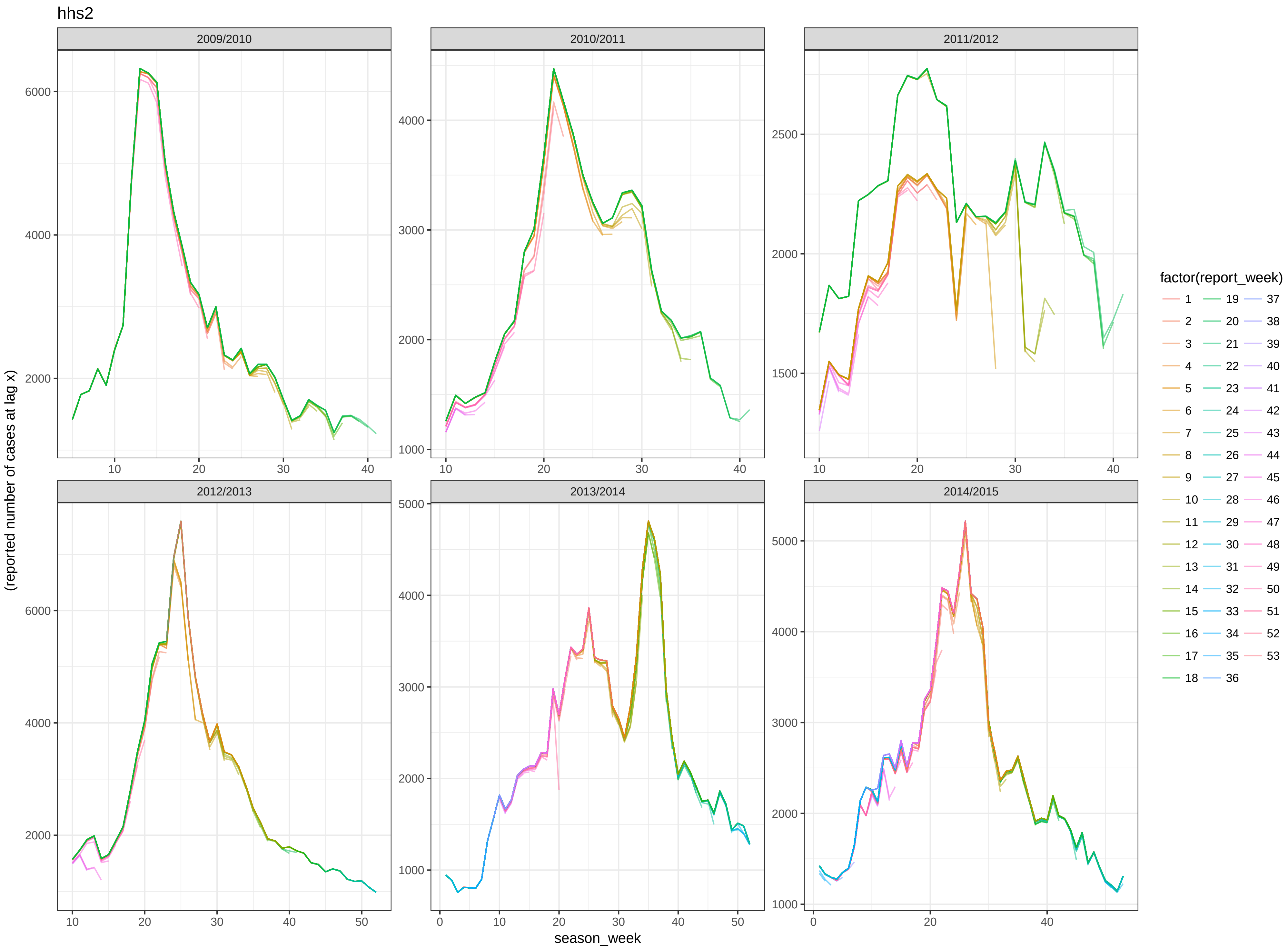

hhs10
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x) /
(reported number of cases at lag 51)
1.00
0.75
0.50
lag
0
10
20
30
40
50
season_week
50
40
30
20
10

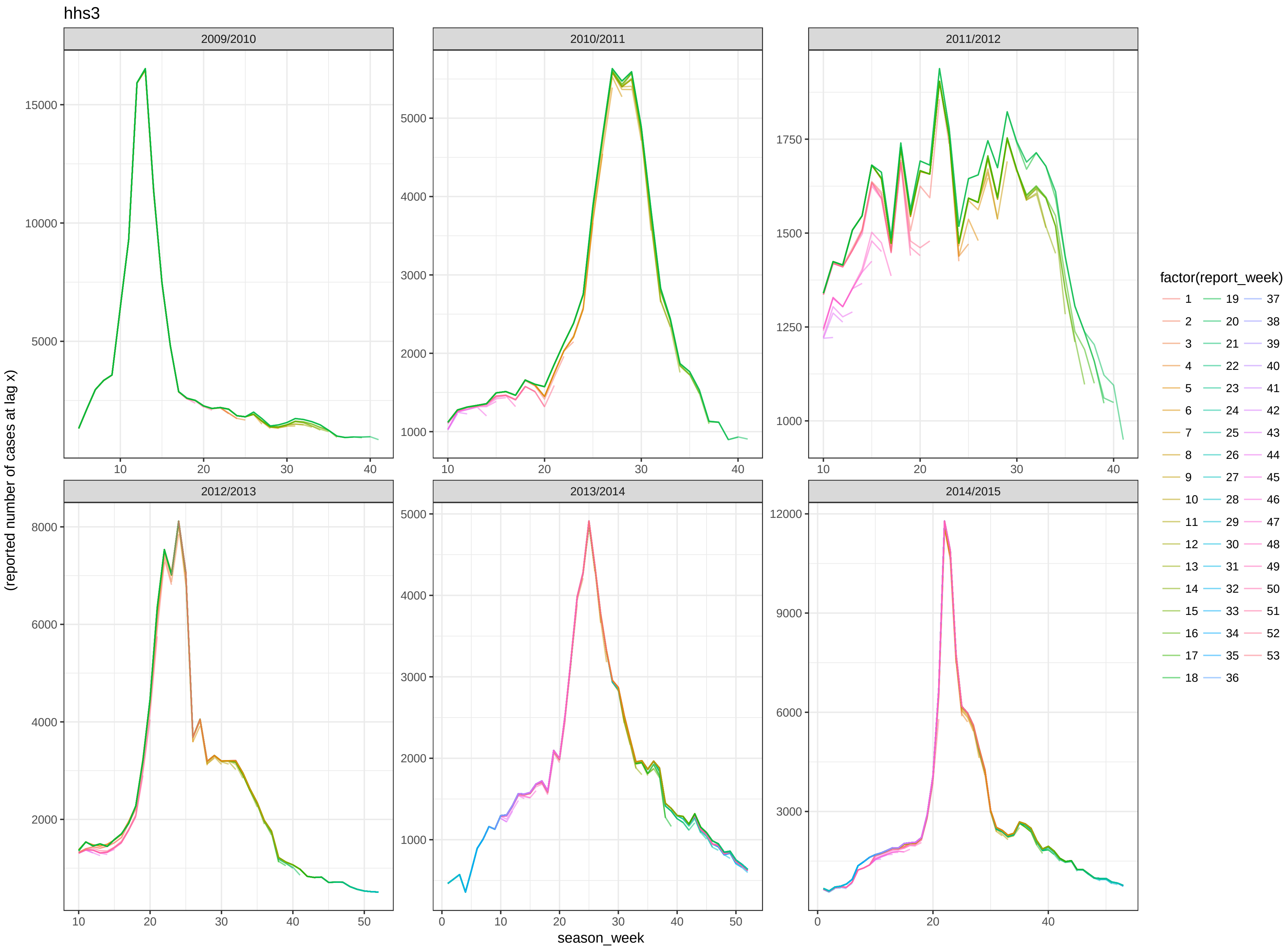
0.2 Visualizing backfill by region and season – Total by report date

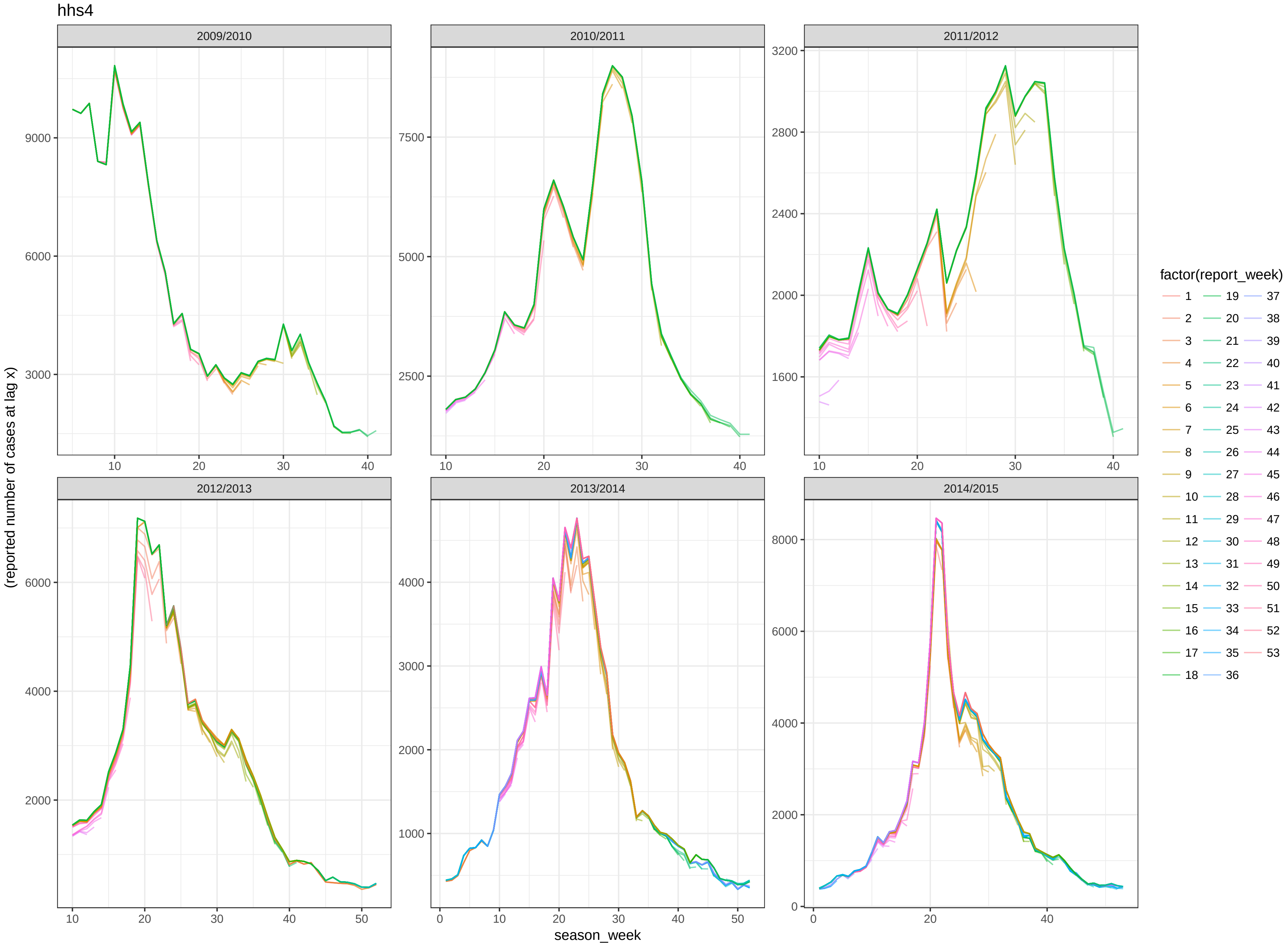
```r
seasons_to_plot <- paste0(2003:2014, "/", 2004:2015)

for(region_val in unique(flu$region)) {
  p <- ggplot(data = flu %>% filter(region == region_val & season %in% seasons_to_plot)) +
    geom_line(
      aes(x = season_week,
        y = num_ili,
        group = factor(report_week_date),
        color = factor(report_week)),
      alpha = 0.5) +
#    scale_color_gradientn(colors = rainbow(7)) +
    facet_wrap(˜ season, scales = "free") +
    ylab("(reported number of cases at lag x)") +
    ggtitle(region_val) +
    theme_bw()
  print(p)
}
```

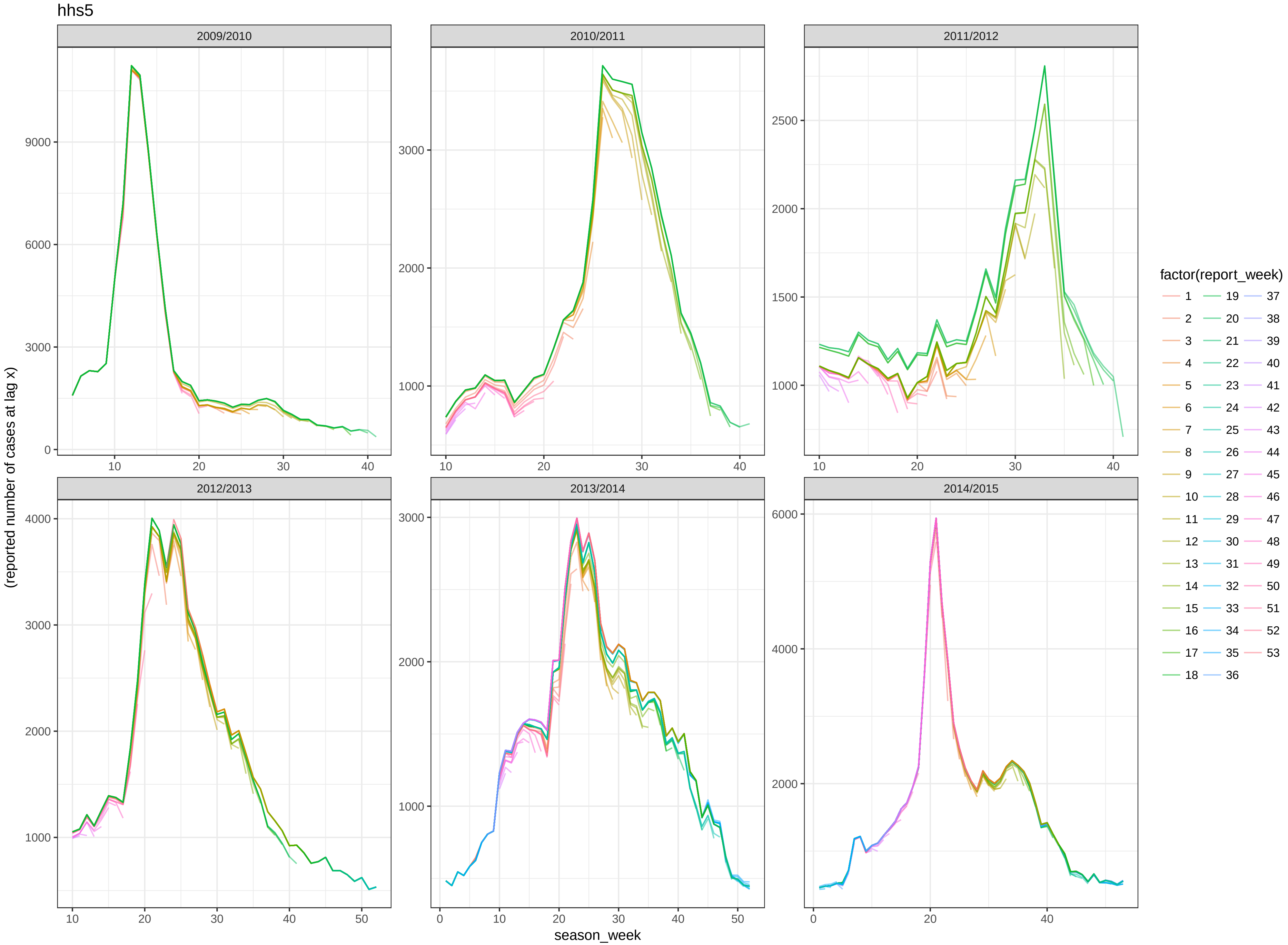

nat
2003/2004
2004/2005
2005/2006
2006/2007
2007/2008
2008/2009
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
0
5000
10000
15000
20000
25000
5000
10000
15000
20000
3000
6000
9000
12000
5000
10000
30000
20000
10000
40000
60000
20000
30000
10
20
30
40
50
(reported number of cases at lag x)
season_week
factor(report_week)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

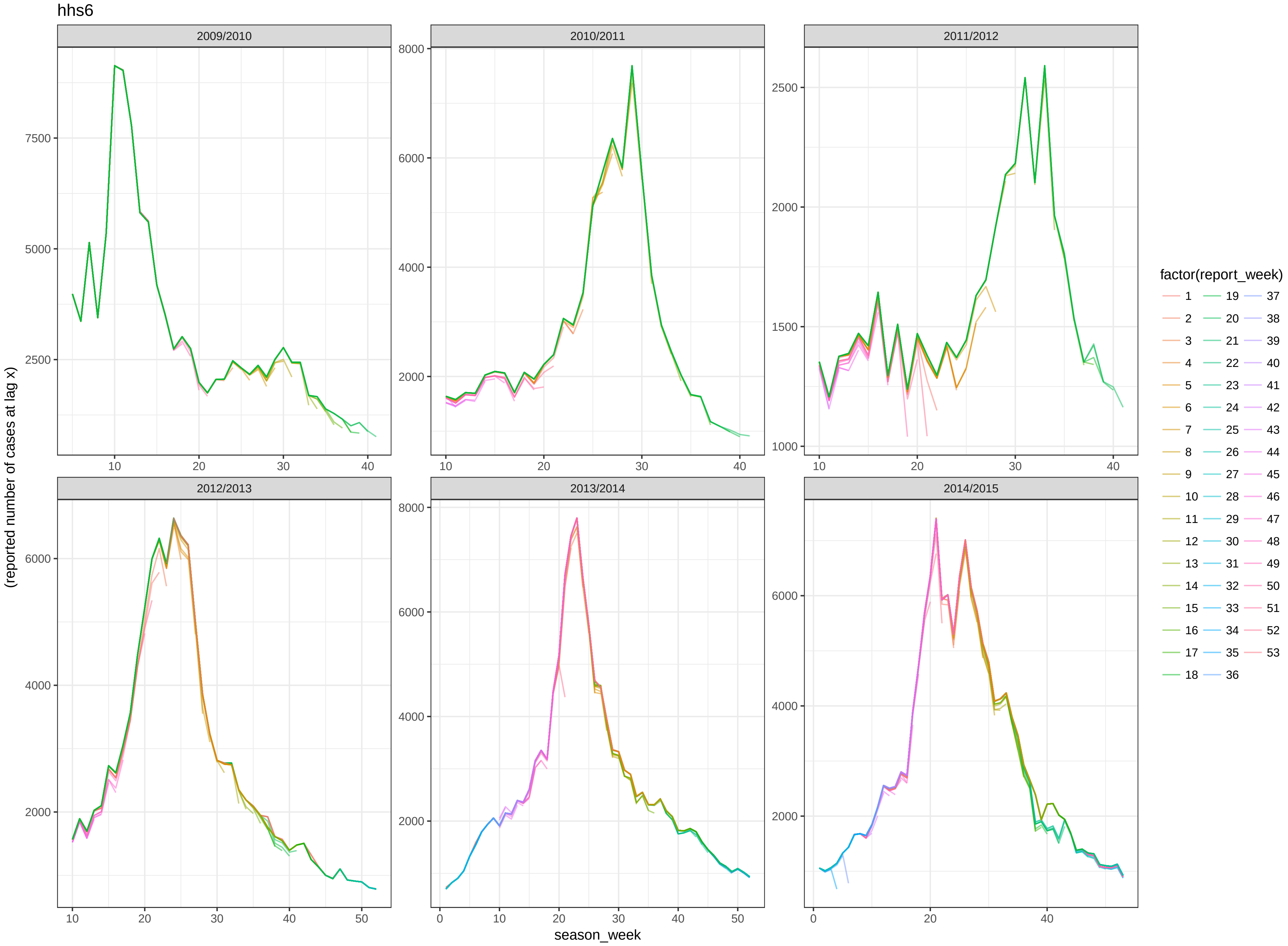

hhs1
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x)
season_week
0
1000
2000
3000
4000
500
1000
200
300
400
500
500
1000
1500
250
500
750
1000
500
1000
1500
2000
10
20
30
40
50
0
factor(report_week)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53

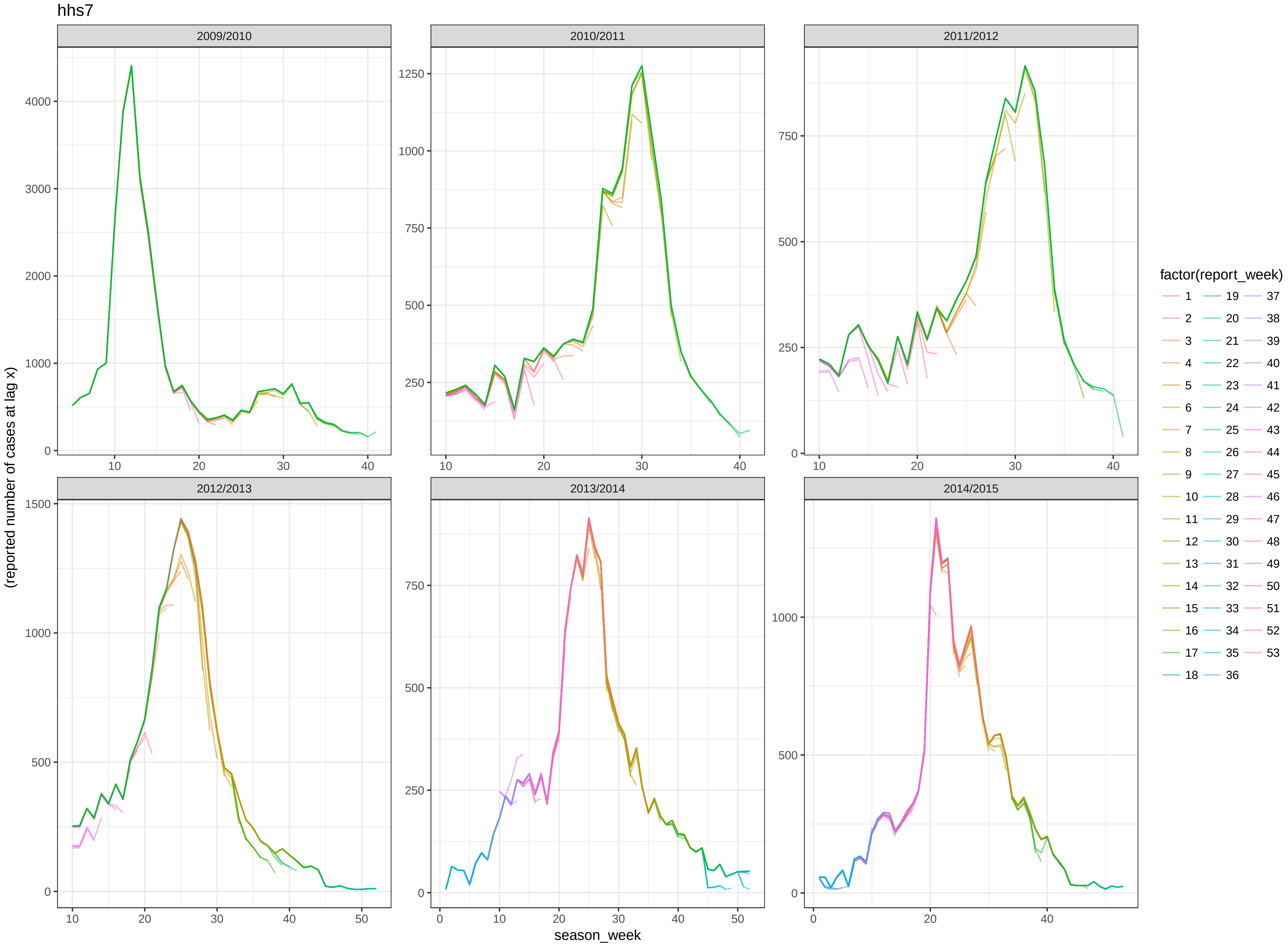

hhs2
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
6000
4000
2000
4000
3000
2000
1000
2500
2000
1500
6000
4000
2000
5000
4000
3000
2000
1000
5000
4000
3000
2000
1000
10
20
30
40
50
0
(reported number of cases at lag x)
season_week
factor(report_week)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53


hhs3
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
15000
10000
5000
10
20
30
40
5000
4000
3000
2000
1000
10
20
30
40
1750
1500
1250
1000
10
20
30
40
8000
6000
4000
2000
10
20
30
40
50
5000
4000
3000
2000
1000
0
10
20
30
40
50
12000
9000
6000
3000
0
20
40
(reported number of cases at lag x)
season_week
factor(report_week)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

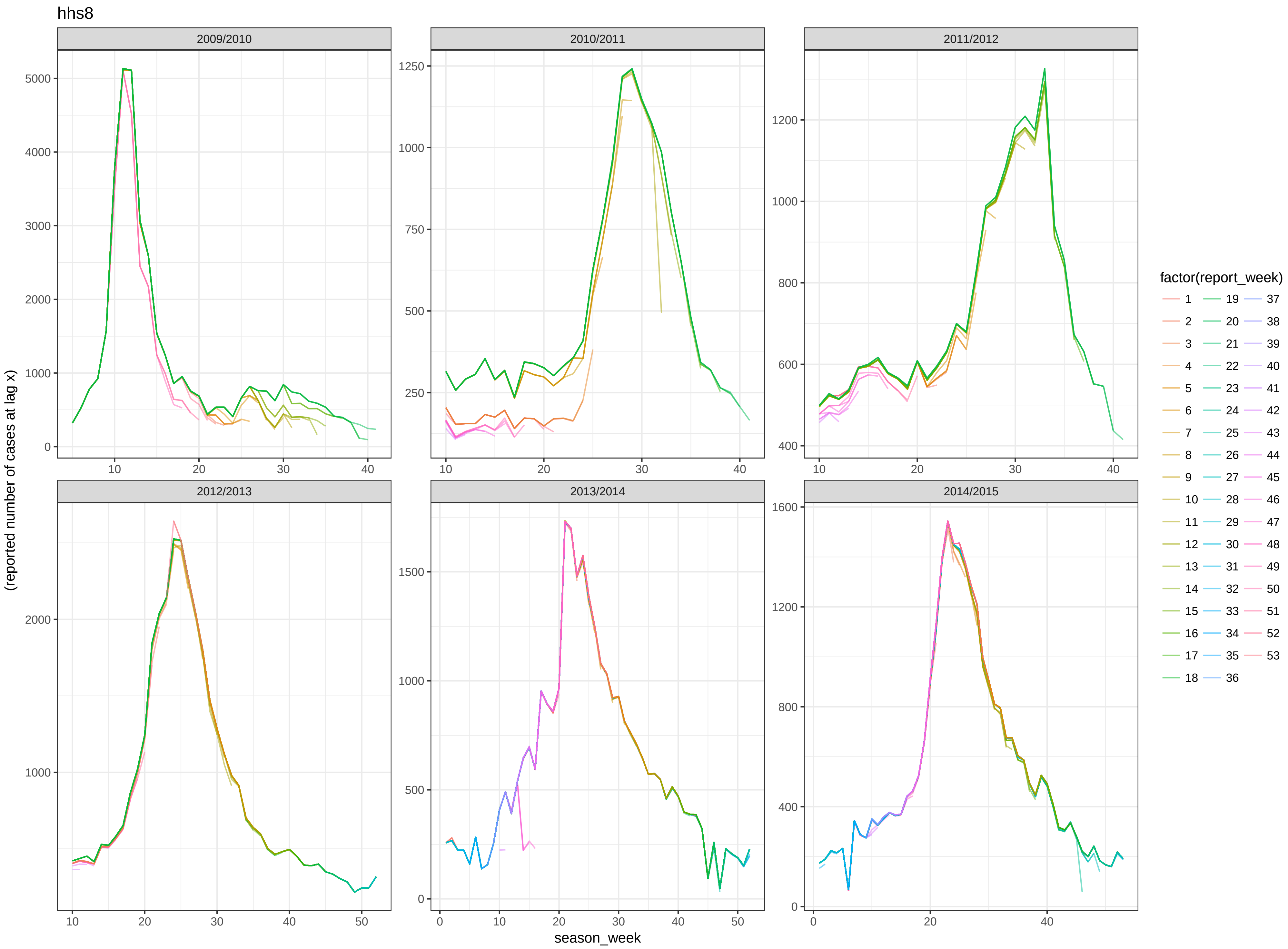

hhs4
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
9000
6000
3000
7500
5000
2500
3200
2800
2400
2000
1600
10
20
30
40
6000
4000
2000
10
20
30
40
50
4000
3000
2000
1000
0
10
20
30
40
50
8000
6000
4000
2000
0
0
20
40
(reported number of cases at lag x)
season_week
factor(report_week)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

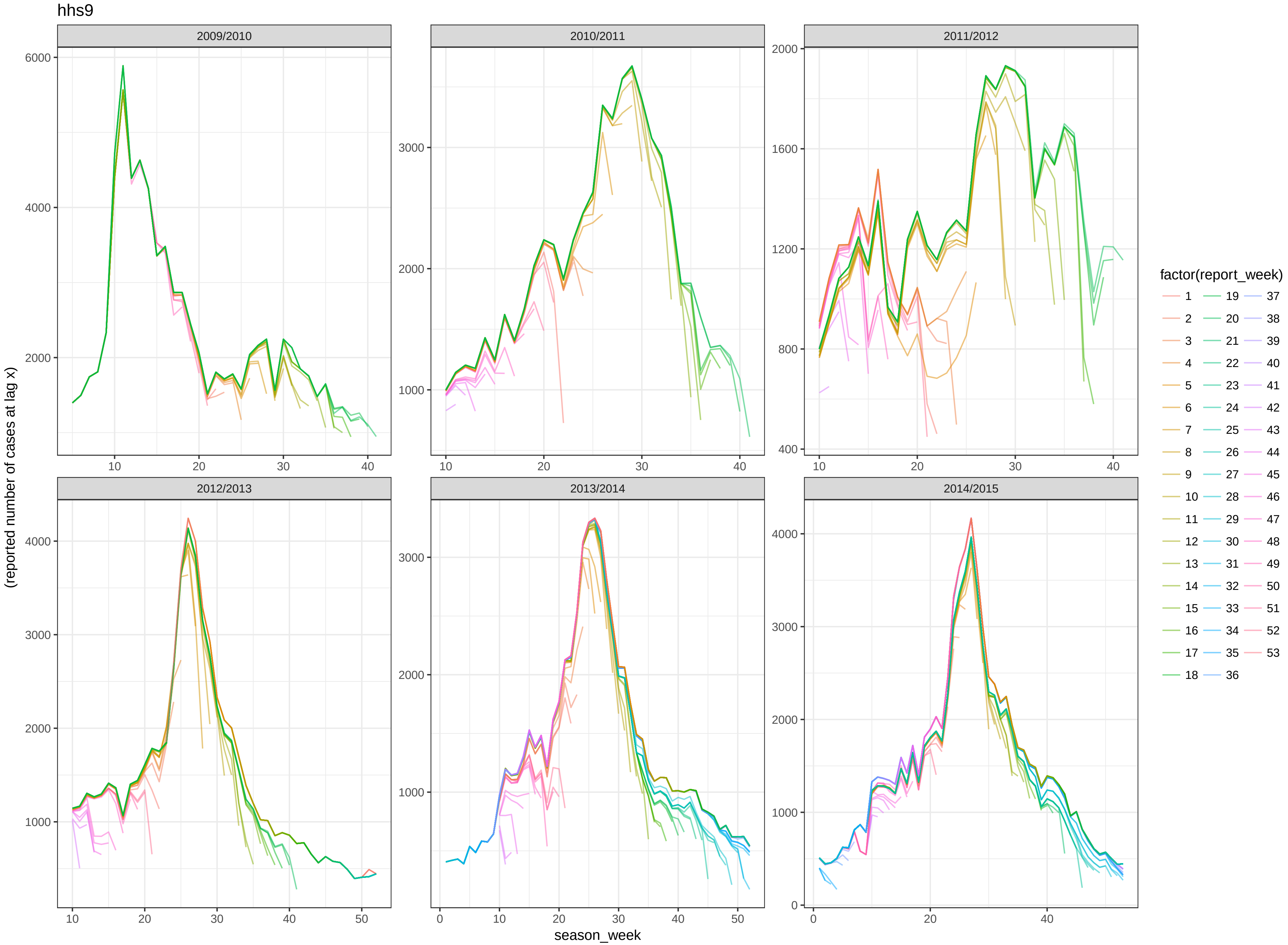

hhs5
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x)
season_week
0
3000
6000
9000
1000
2000
3000
1000
1500
2000
2500
1000
2000
3000
4000
1000
2000
3000
2000
4000
6000
0
10
20
30
40
50
factor(report_week)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

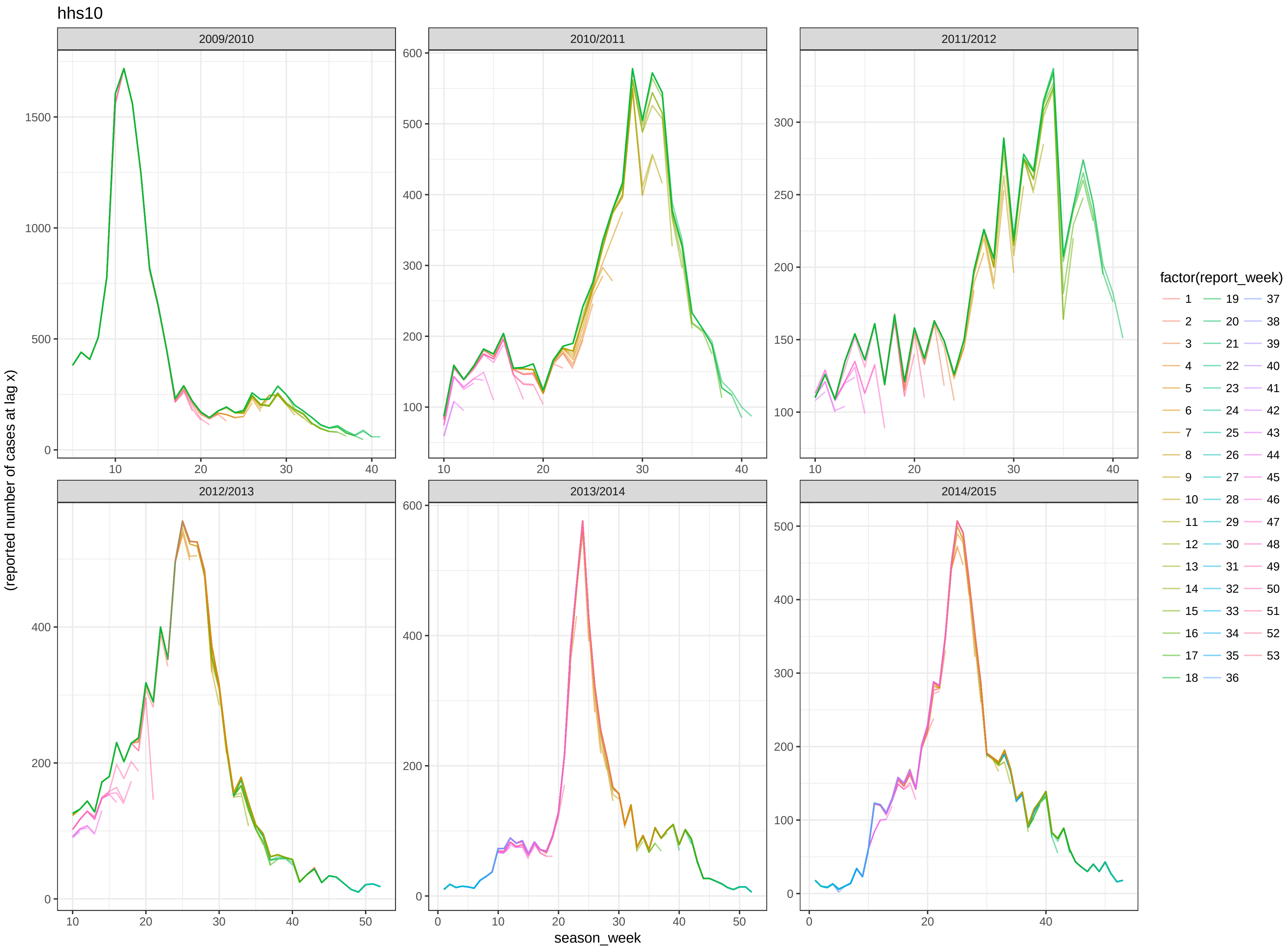

hhs6
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
7500
5000
2500
8000
6000
4000
2000
2500
2000
1500
1000
10
20
30
40
50
0
(reported number of cases at lag x)
season_week
factor(report_week)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53



hhs7
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x)
season_week
0
1000
2000
3000
4000
250
500
750
1000
1250
0
250
500
750
0
500
1000
1500
0
250
500
750
0
500
1000
10
20
30
40
50
0
factor(report_week)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

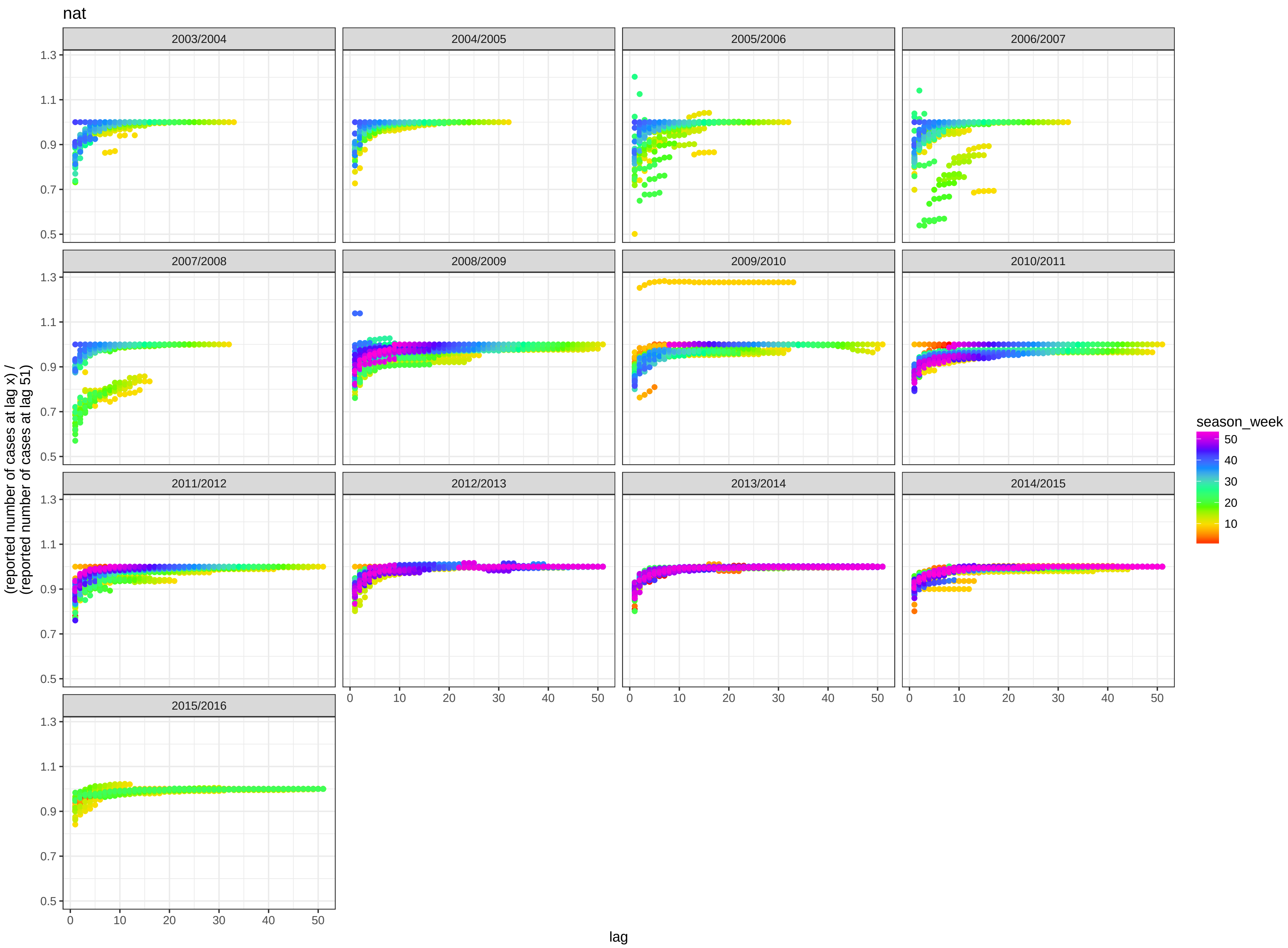

hhs8
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x)
season_week
0
1000
2000
3000
4000
5000
250
500
750
1000
1250
400
600
800
1000
1200
2000
1000
0
500
1000
1500
0
400
800
1200
1600
10
20
30
40
50
factor(report_week)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

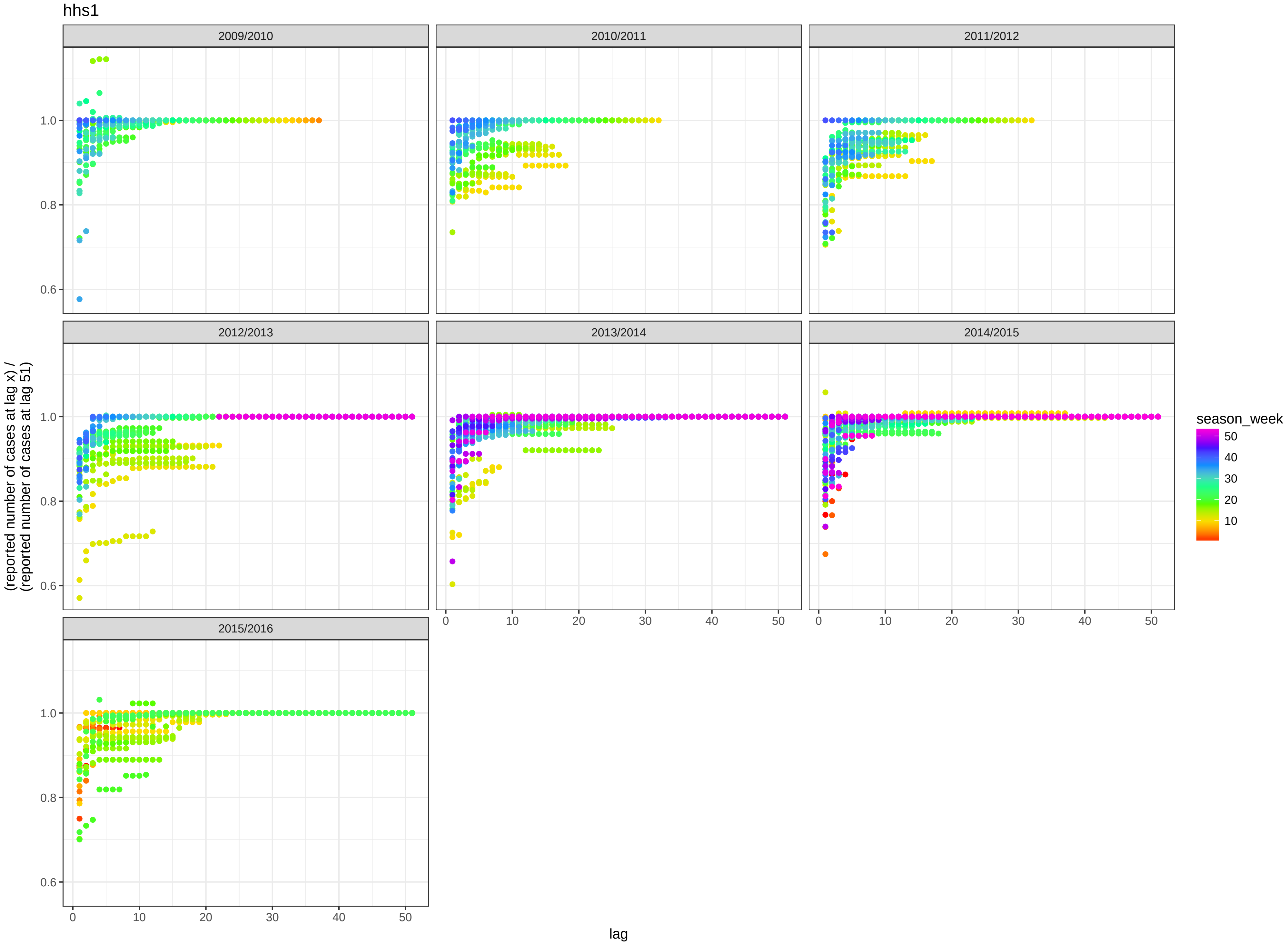

hhs9
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
6000
4000
2000
3000
2000
1000
2000
1600
1200
800
400
4000
3000
2000
1000
3000
2000
1000
4000
3000
2000
1000
0
10
20
30
40
50
(reported number of cases at lag x)
season_week
factor(report_week)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

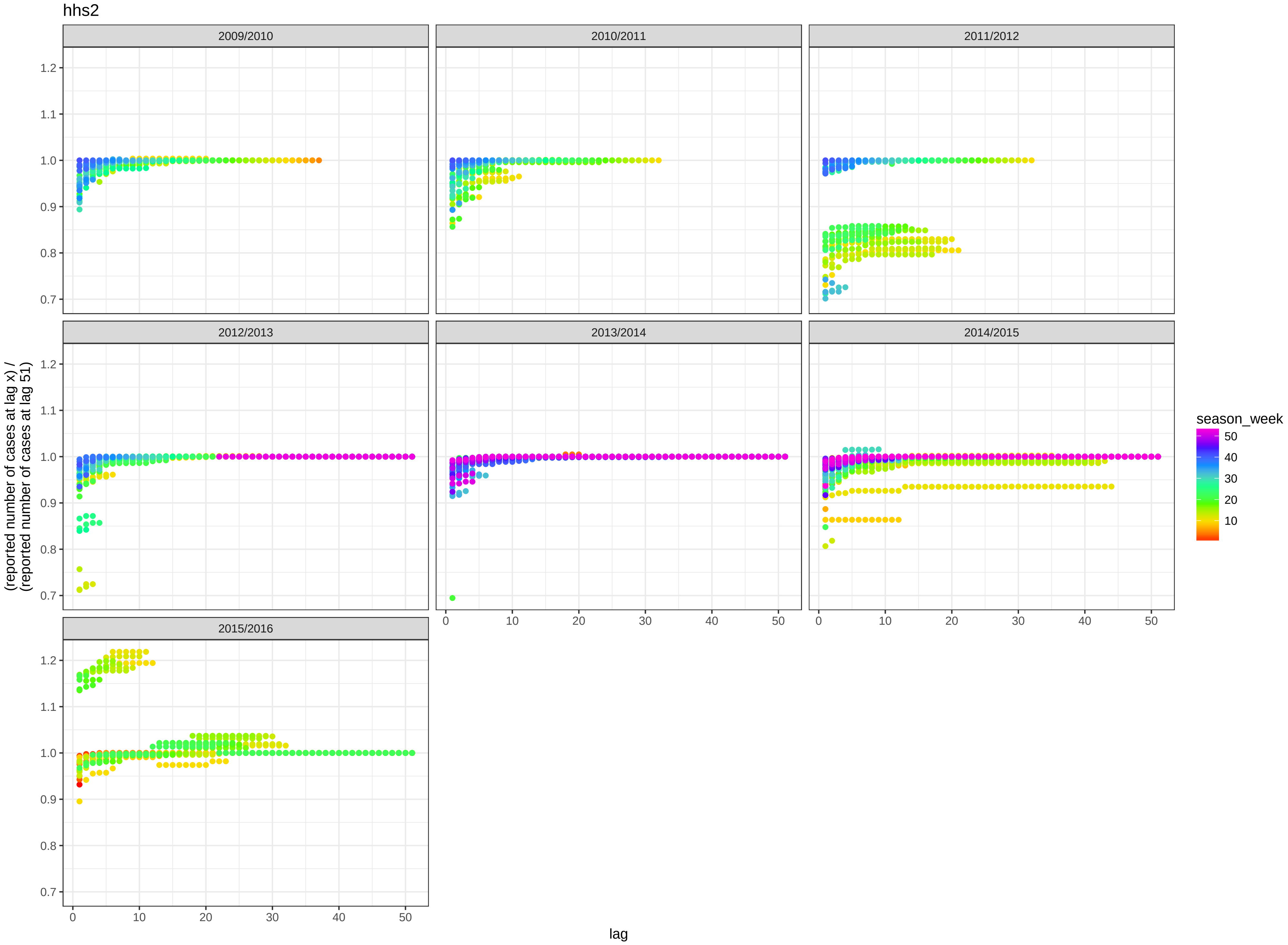

hhs10
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
(reported number of cases at lag x)
season_week
factor(report_week)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53
0 500 1000 1500
100 200 300 400 500 600
100 150 200 250 300
0 200 400
0 200 400 600
0 100 200 300 400 500
10 20 30 40
10 20 30 40 50
0 10 20 30 40 50
0 20 40

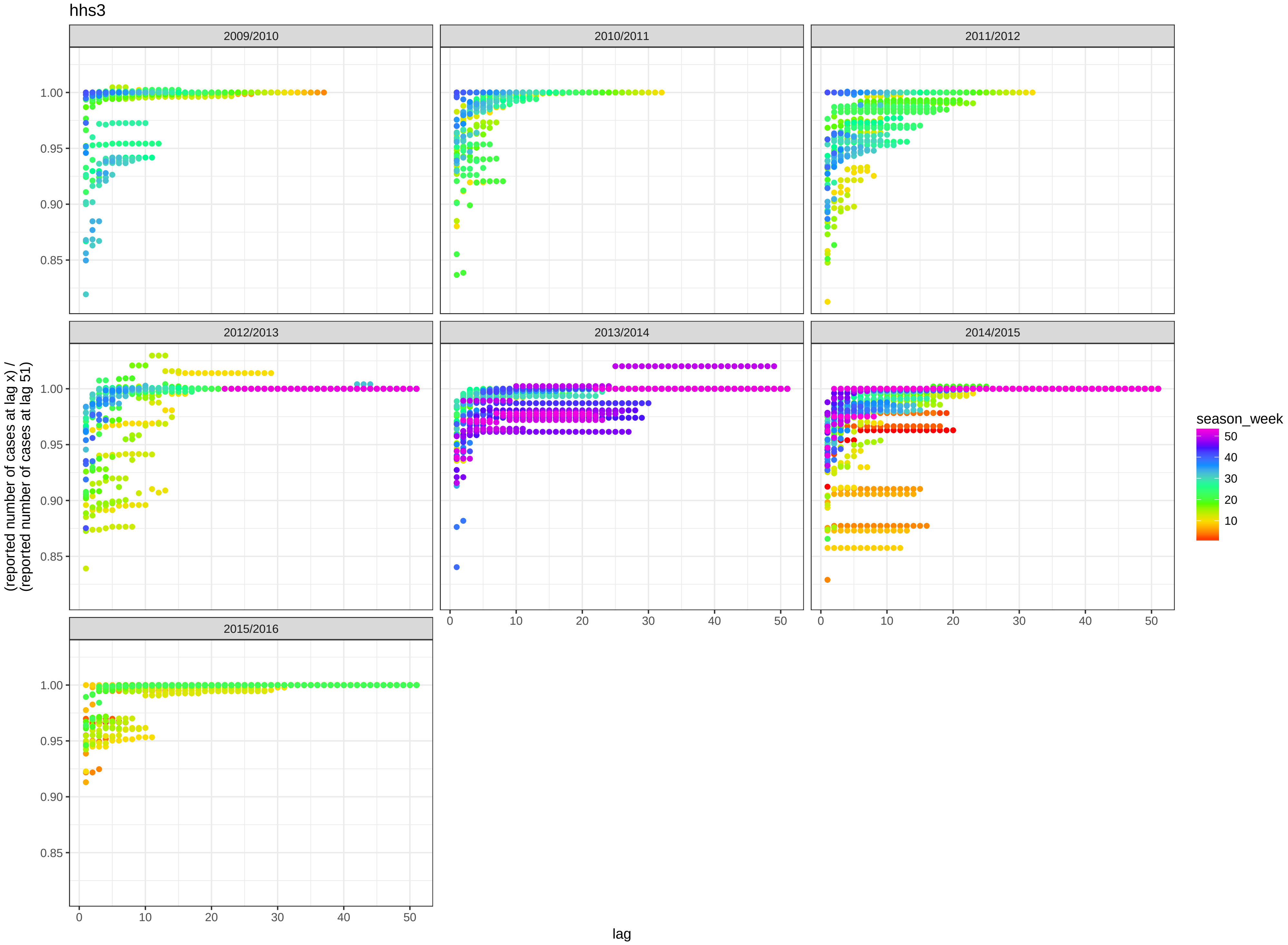
0.3 Visualizing backfill by region and season – ratios heatmap

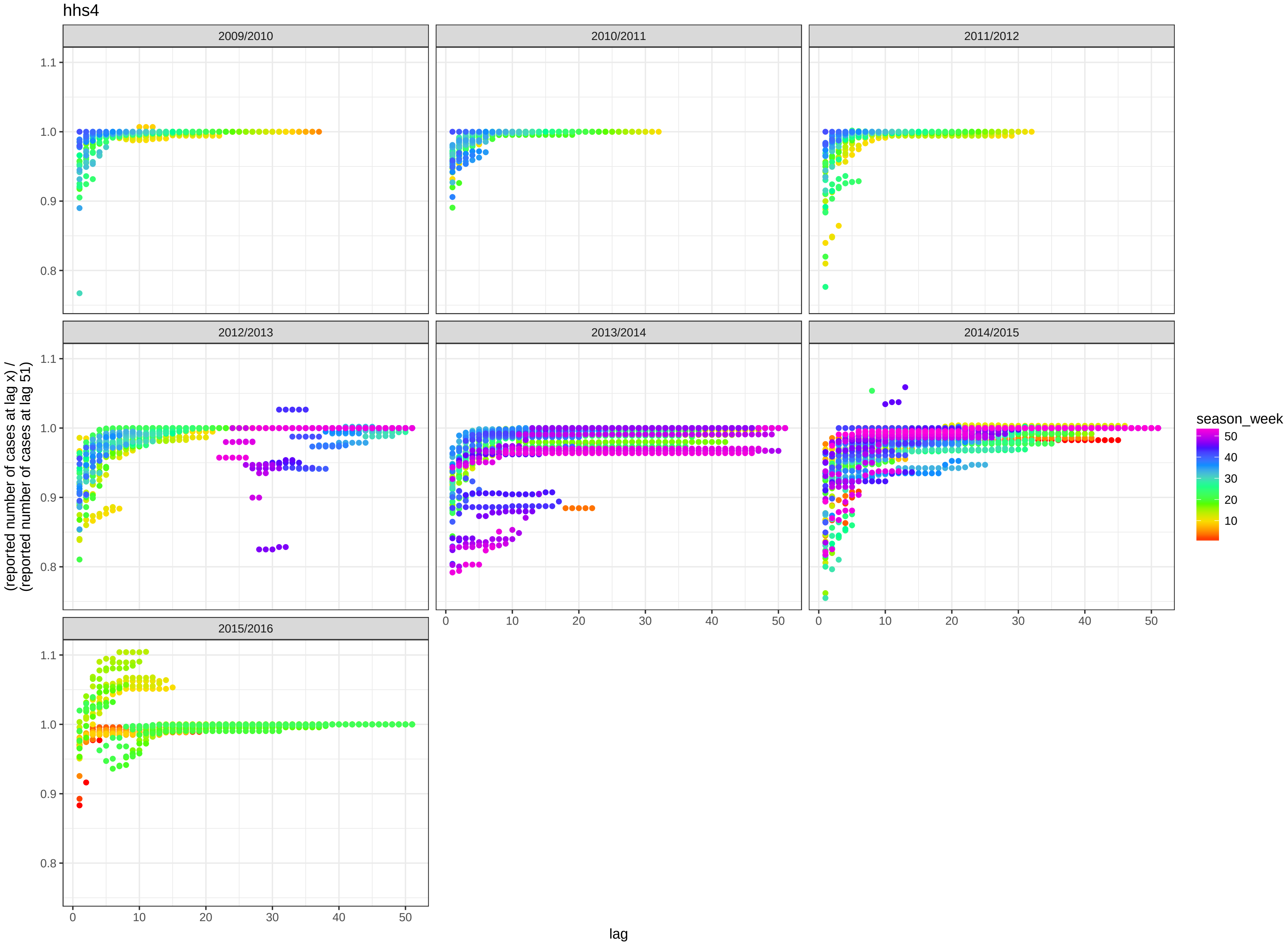
```r
seasons_to_plot <- paste0(2003:2014, "/", 2004:2015)

for(region_val in unique(flu$region)) {
  p <- ggplot() +
    geom_point(aes(x = lag, y = exp(diff_log_curr_final_num_ili), colour = season_week),
      data = flu %>% filter(region == region_val)) +
      scale_color_gradientn(colors = rainbow(7)) +
    facet_wrap( ~ season) +
    ggtitle(region_val) +
    ylab("(reported number of cases at lag x) /\n(reported number of cases at lag 51)") +
    theme_bw()
  print(p)
}
```

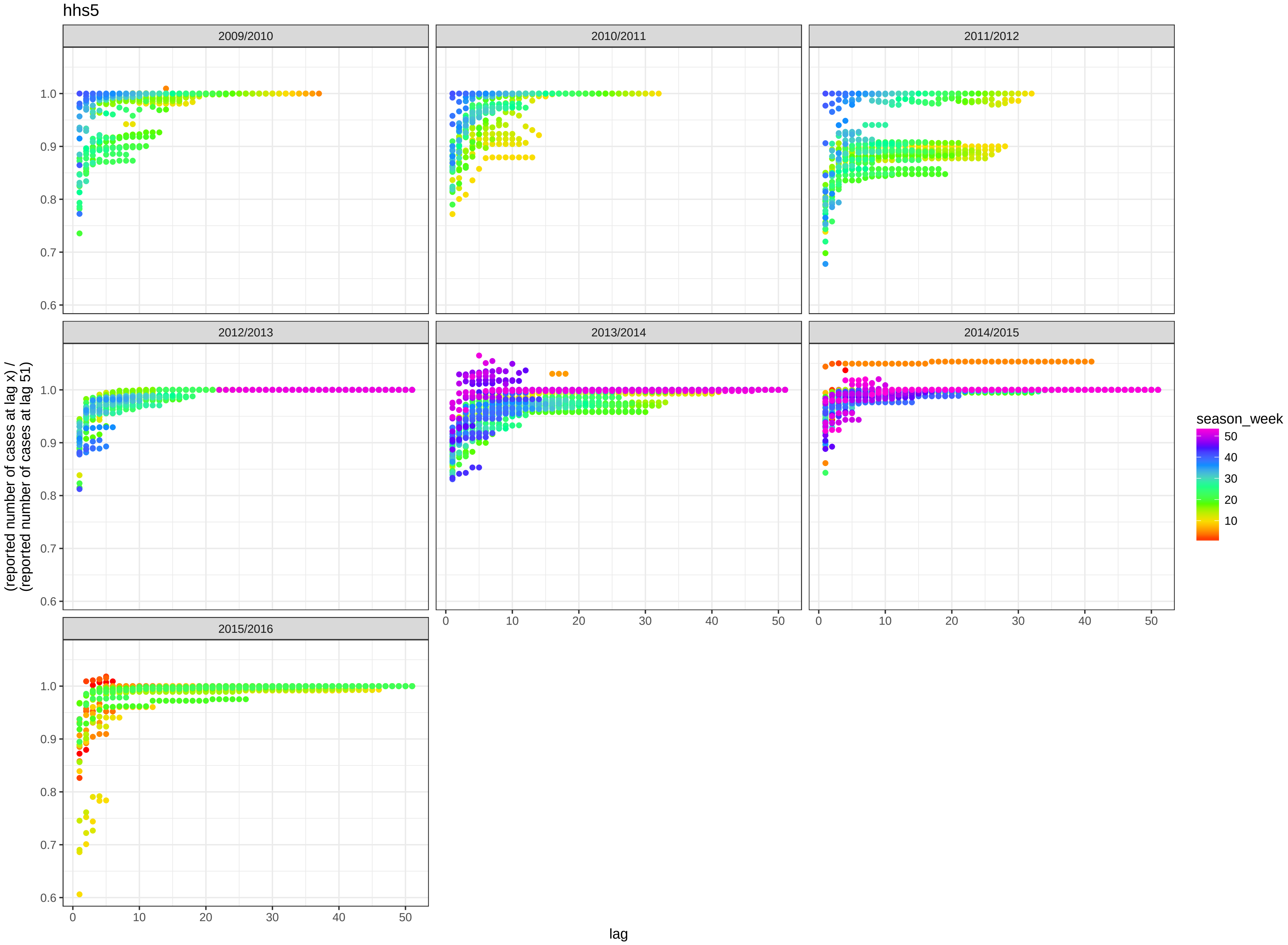

nat
2003/2004
2004/2005
2005/2006
2006/2007
2007/2008
2008/2009
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
1.3
1.1
0.9
0.7
0.5
0
10
20
30
40
50
(reported number of cases at lag x) /
(reported number of cases at lag 51)
lag
season_week
50
40
30
20
10

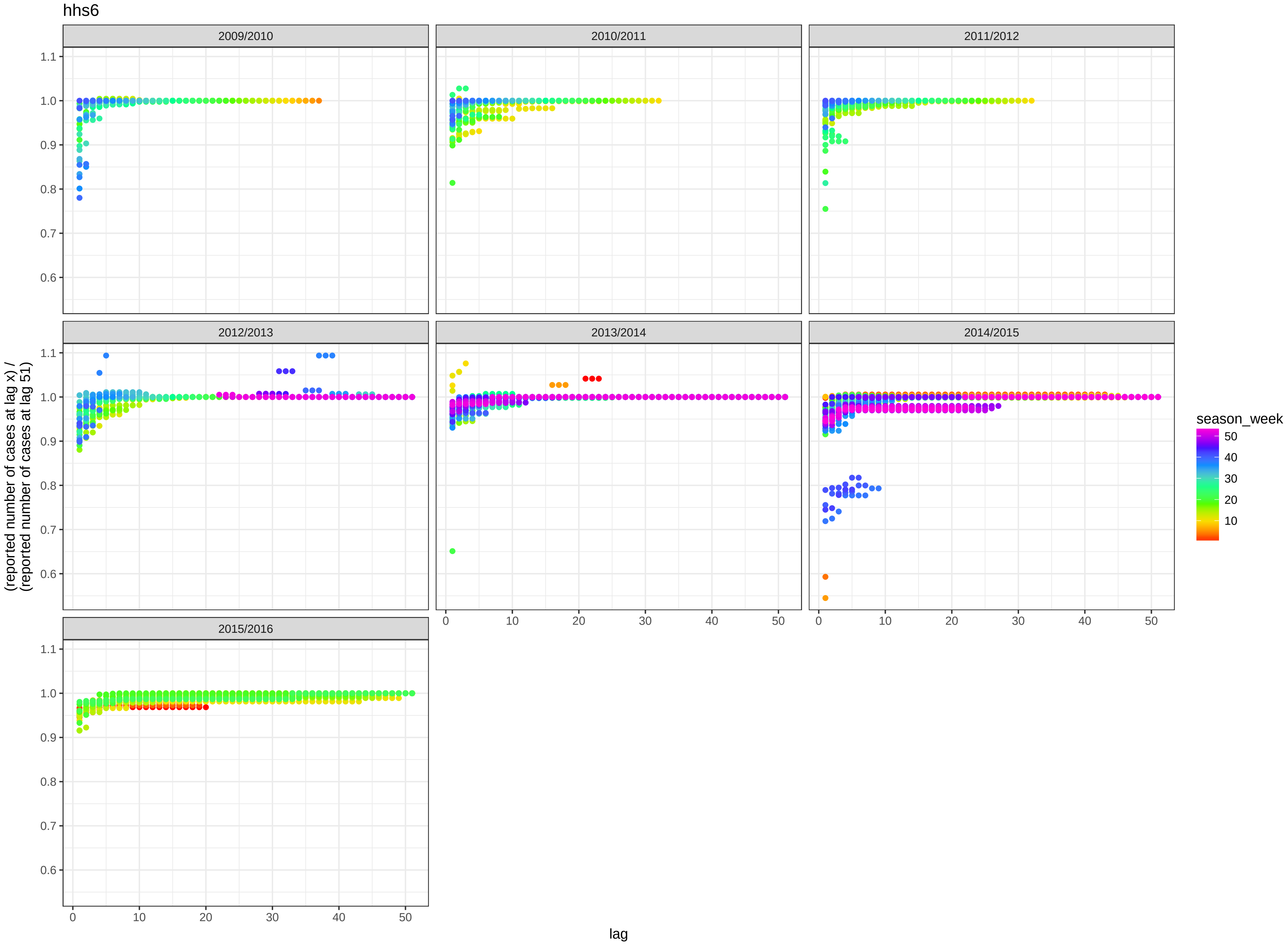

hhs1
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
(reported number of cases at lag x) /
(reported number of cases at lag 51)
0.6
0.8
1.0
lag
0
10
20
30
40
50
season_week
50
40
30
20
10

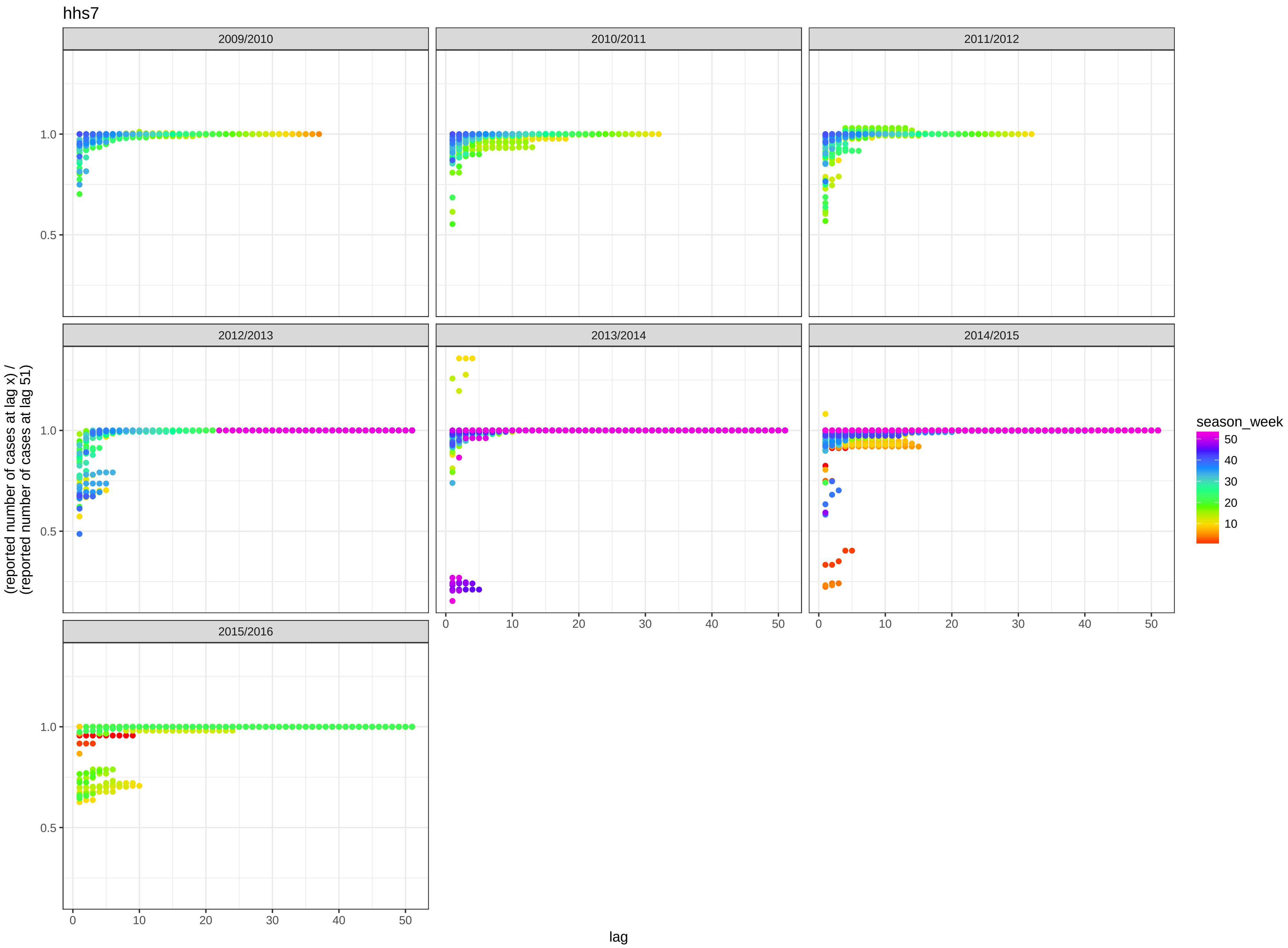

hhs2
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
1.2
1.1
1.0
0.9
0.8
0.7
0
10
20
30
40
50
(reported number of cases at lag x) /
(reported number of cases at lag 51)
lag
season_week
50
40
30
20
10

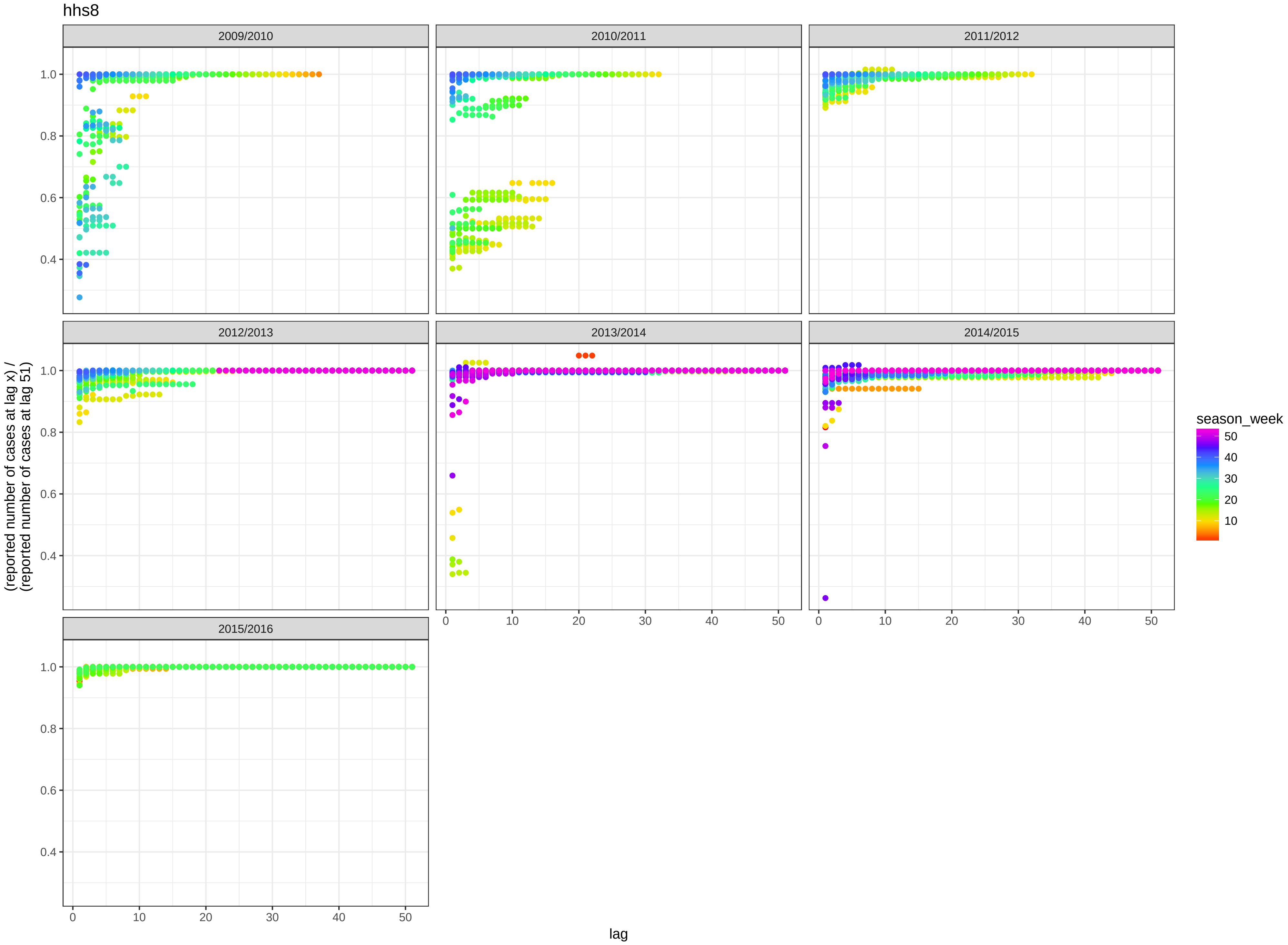

hhs3
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
1.00
0.95
0.90
0.85
0
10
20
30
40
50
(reported number of cases at lag x) /
(reported number of cases at lag 51)
lag
season_week
50
40
30
20
10

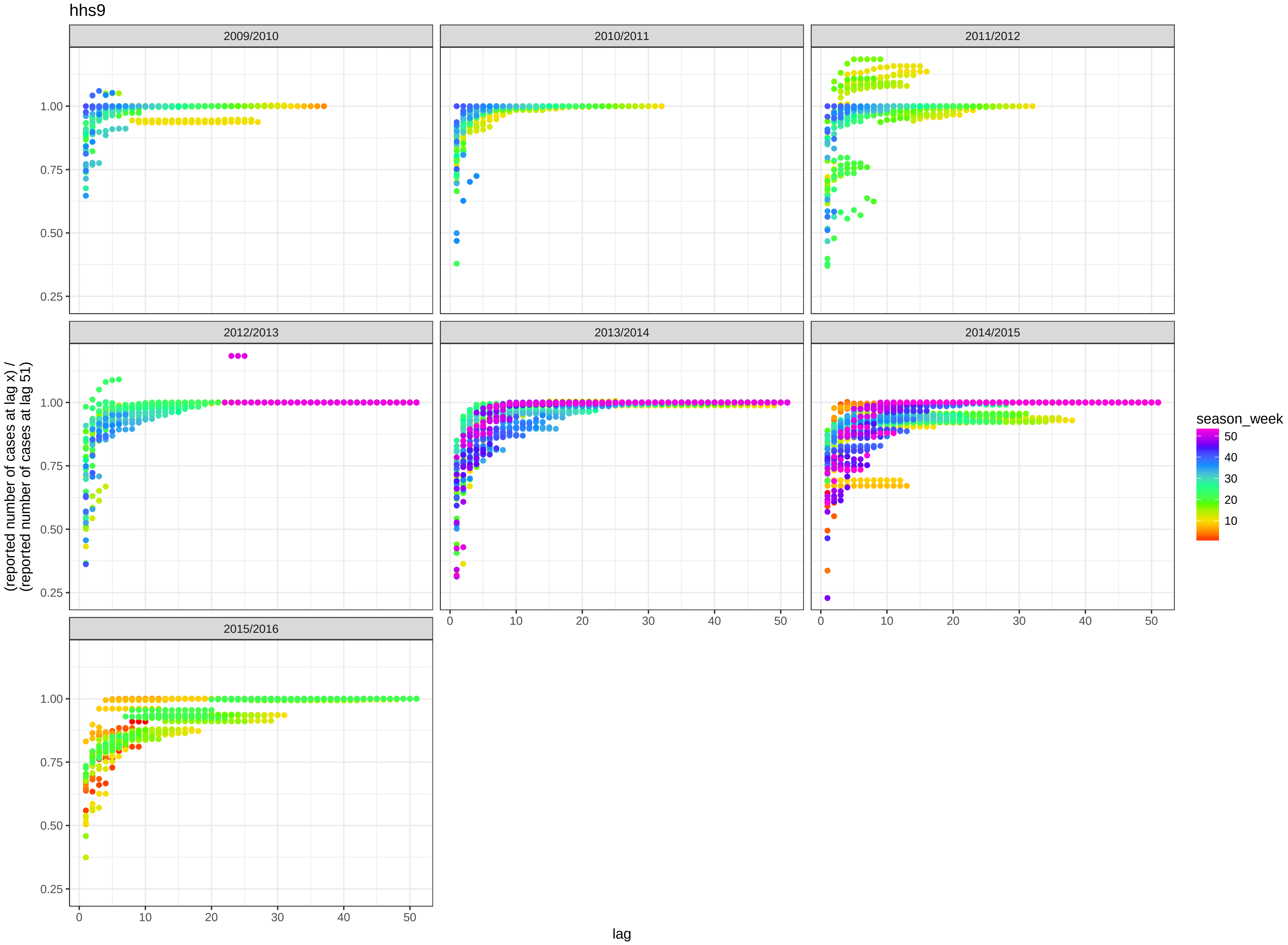

hhs4
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
1.1
1.0
0.9
0.8
(reported number of cases at lag x) /
(reported number of cases at lag 51)
0
10
20
30
40
50
lag
season_week
50
40
30
20
10

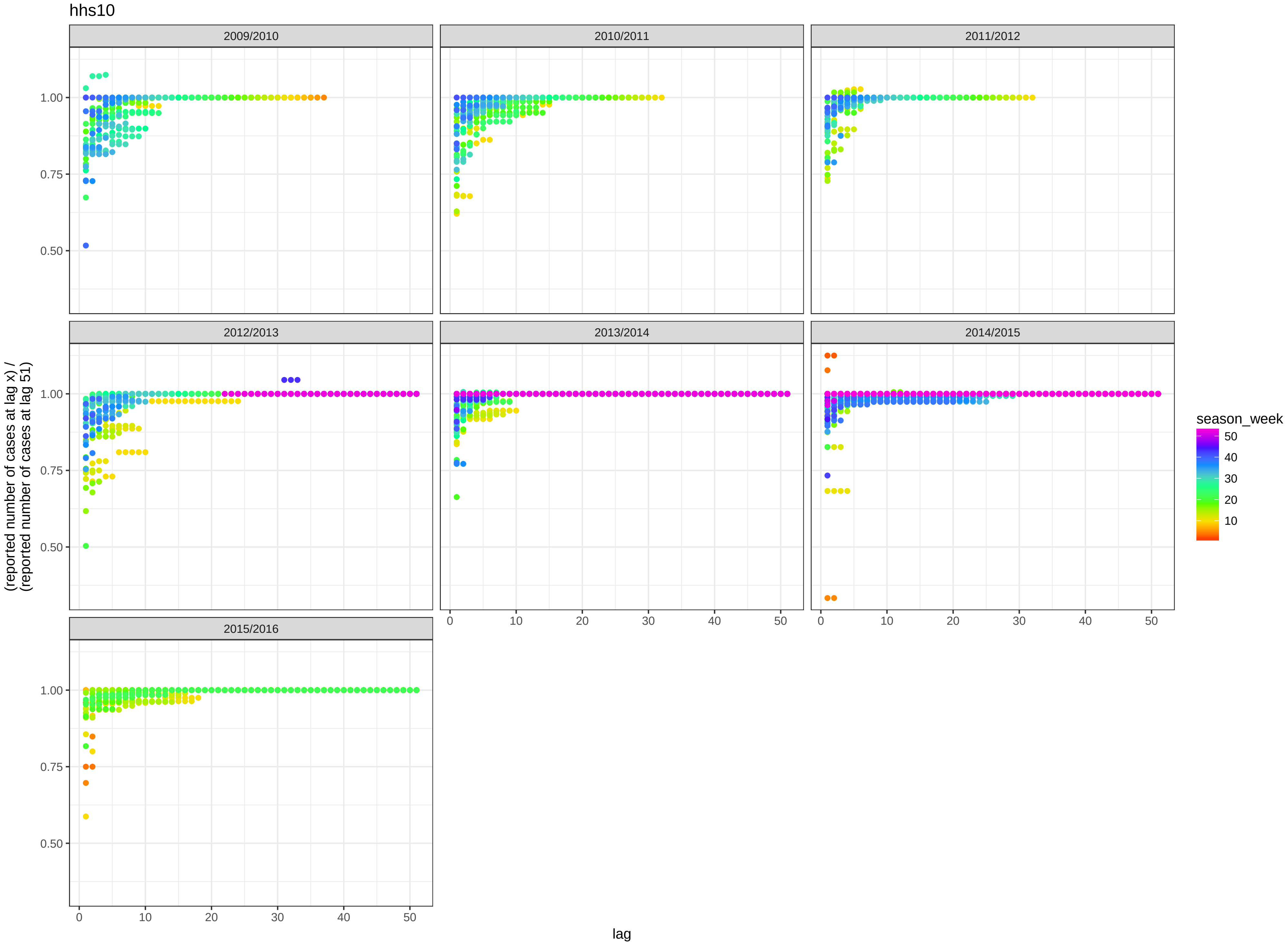

hhs5
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
1.0
0.9
0.8
0.7
0.6
0
10
20
30
40
50
(reported number of cases at lag x) /
(reported number of cases at lag 51)
lag
season_week
50
40
30
20
10


hhs6
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
1.1
1.0
0.9
0.8
0.7
0.6
0
10
20
30
40
50
(reported number of cases at lag x) /
(reported number of cases at lag 51)
lag
season_week
50
40
30
20
10


hhs7
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
1.0
0.5
0
10
20
30
40
50
lag
(reported number of cases at lag x) /
(reported number of cases at lag 51)
season_week
50
40
30
20
10

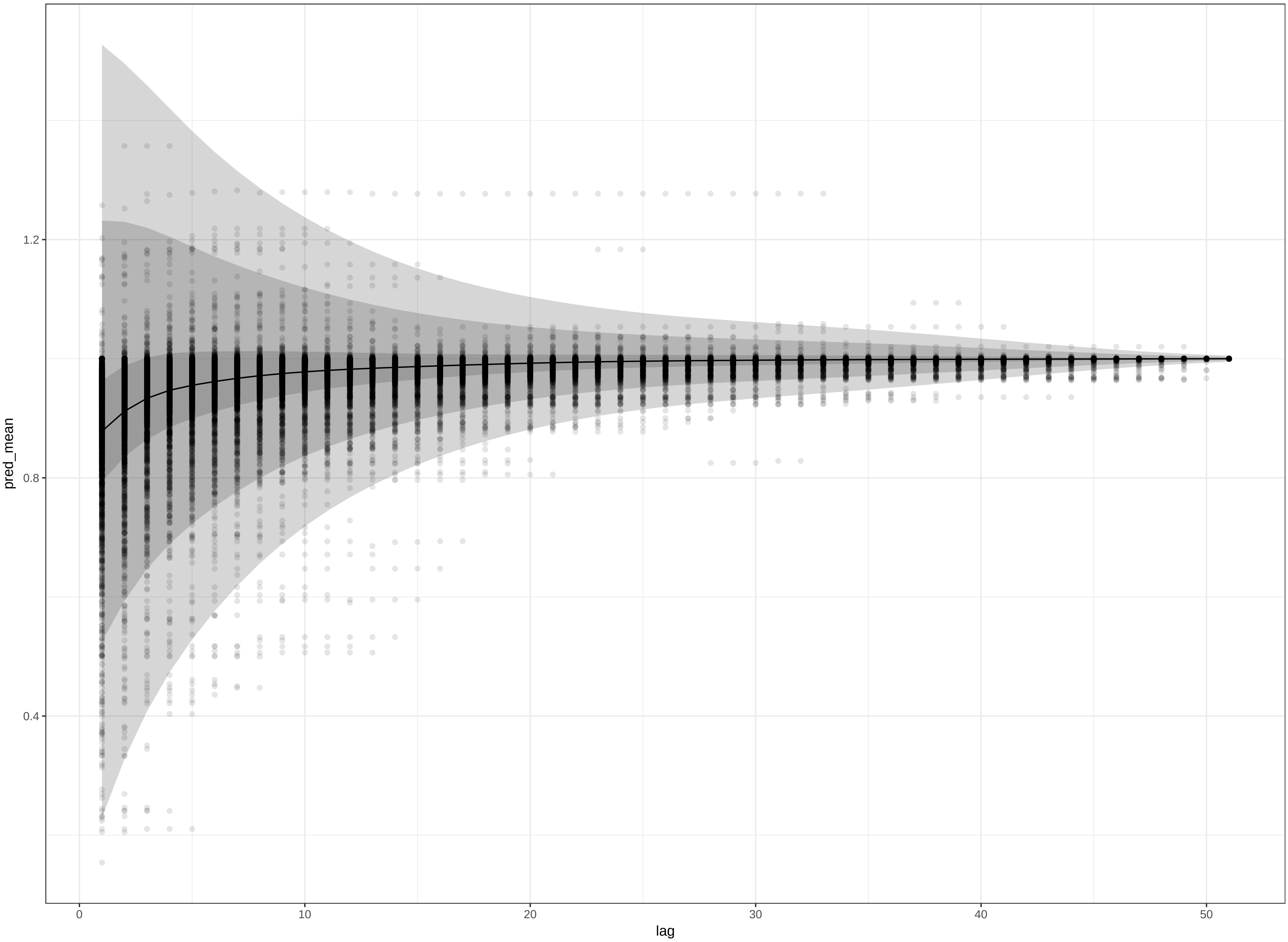

hhs8
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
1.0
0.8
0.6
0.4
0
10
20
30
40
50
(reported number of cases at lag x) /
(reported number of cases at lag 51)
lag
season_week
50
40
30
20
10


hhs9
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
(reported number of cases at lag x) /
(reported number of cases at lag 51)
1.00
0.75
0.50
0.25
0
10
20
30
40
50
lag
season_week
50
40
30
20
10


hhs10
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
(reported number of cases at lag x) /
(reported number of cases at lag 51)
1.00
0.75
0.50
0
10
20
30
40
50
lag
season_week
50
40
30
20
10

0.4 Model 1

Combining all regions and seasons, mean specification using a spline with 15 df, variance specification using a spline with 5 df, t_3 distributed errors. Degrees of freedom for splines and t distribution chosen "by eye". Normal distribution used in estimation, t just used to make the plot. Plot shows 50%, 95%, and 99% predictive intervals from the model fit.



pred_mean
1.2
0.8
0.4
0
10
20
30
40
50
lag