TS_Project_HTS

Chinta

2023-03-03

0. Import dependencies.

1. Load the chickenpox data.

Load the data, get the region-county data, join the chickenpox cases with region-county data, and make it a monthly structure.

```
hch <- read.csv("hungary_chickenpox.csv")  #read the data

hch_r_c <- read.csv("hungary_rgn_county.csv") #get the region-counties info

names(hch_r_c) <- c("region", "rgn","county") #rename the columns

hch.df <- data.frame(hch) #Change to data frame

hch.df <- hch.df  %>%
  pivot_longer (cols =! Date,names_to = "county",
                values_to ="cases_count") #Unpivot the data

hch.df <- hch.df %>% left_join(hch_r_c) #Join the cases data with region-counties data

hch.df <- hch.df %>% select ( Date, region, rgn, county, cases_count) #arrange the columns

hch.df$Date <- as.Date(hch.df$Date, format="%d/%m/%Y") #format the date using as.Date

#hch.df <- hch.df %>% unite ("rgn", rgn:county, sep="") #combine the region and county columns

hch.df <- hch.df %>% select ( Date, region,  county, cases_count) #arrange the columns

hch.df$Date <- yearmonth(hch.df$Date) #format the Date columns to change the date from weekly to monthl

hch.df <- hch.df %>% group_by(Date, region, county) %>% summarise (cases_count = sum(cases_count))
```

2. Create a training and a test tsibble.

Creating the training and test tsibble. Dates less than December 2013 are part of training dataset, while above 12/2013 is test dataset. Additionally, use aggregate_key function to create country-level cases count.

```
hch.tsb <- as_tsibble(hch.df, key=c(region, county), index = Date) #create a tsibble object

#Check for gaps in the data
scan_gaps(hch.tsb) %>%
  count(Date)
```

```
## # A tibble: 0 x 3
## # Groups:   Date, region [0]
## # ... with 3 variables: Date <mth>, region <chr>, n <int>
```

```
hch.tsb_agg <- tsibble(hch.df, key = c(region, county), index = Date) %>%
  aggregate_key(region/county, cases_count = sum(cases_count))

hch.tsb_agg_train <- hch.tsb_agg %>% filter(year(Date) <= 2012)
hch.tsb_agg_test <- hch.tsb_agg %>% filter(year(Date) > 2012)
```

2.1 Process Steps

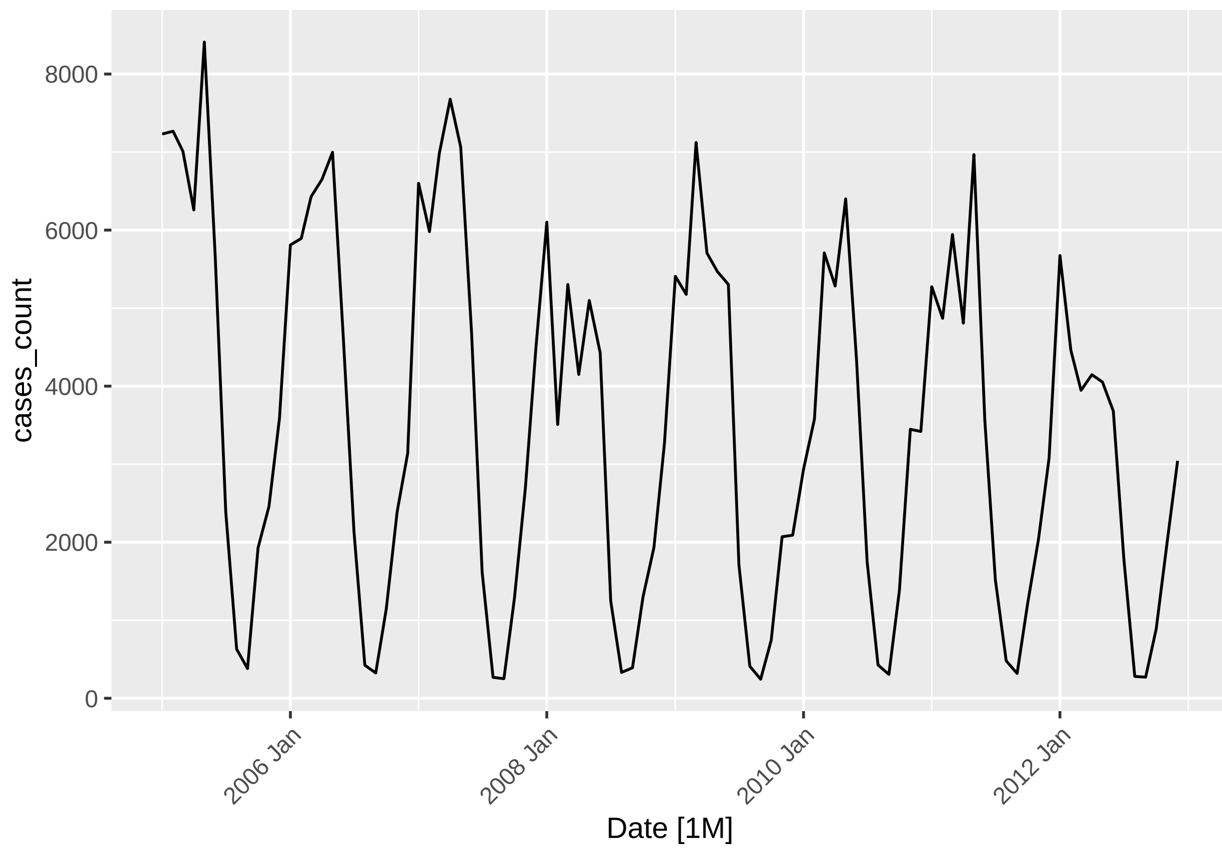
- *data %>% aggregate_key() %>% model() %>% reconcile() %>% forecast()*

3. Plot aggregate (country level) and Central Hungary training data.

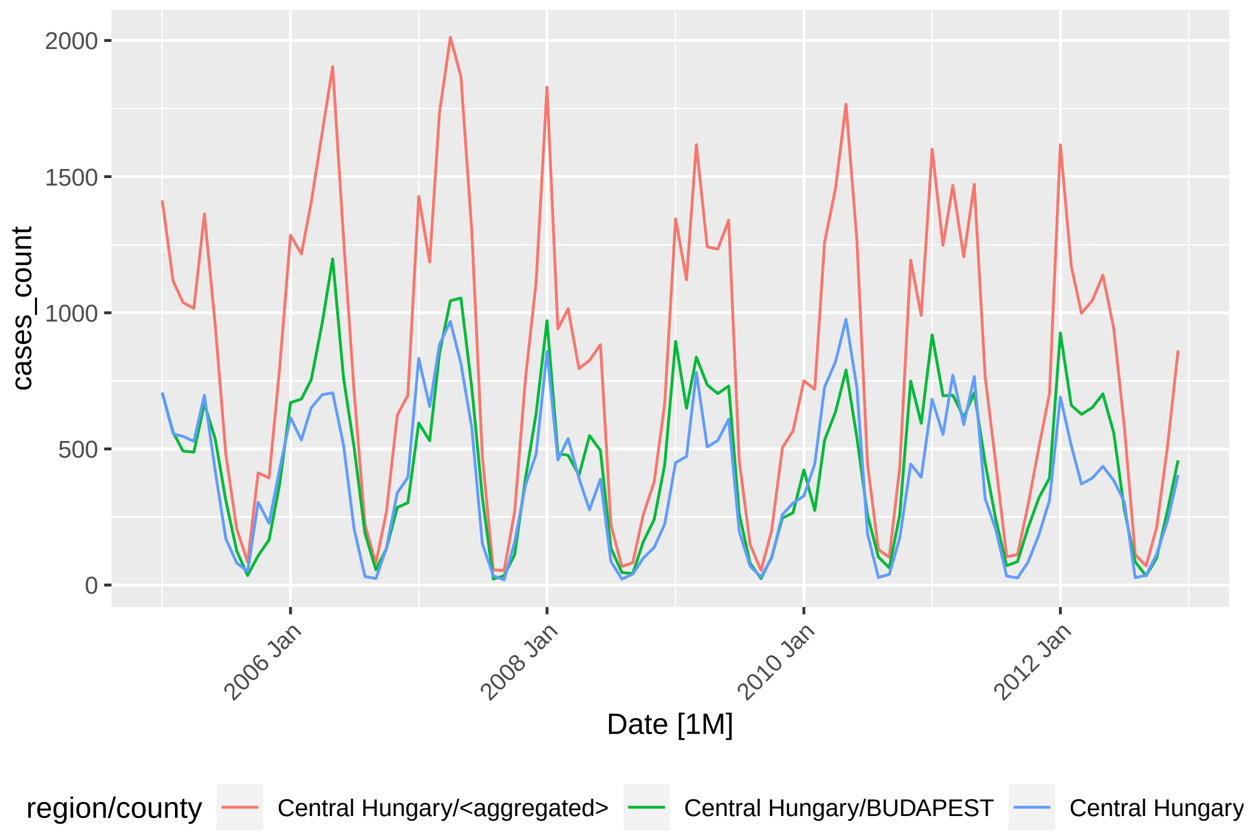
Plot the aggregate and Central Hungary training data to understand data. We see that Central Hungary has pretty heavy influence on the country-level data.

- The variable that we'd like to estimate is the number of cases represented by the 'cases_count' variable. The plot reveals that weak trends and high seasonality are apparent.

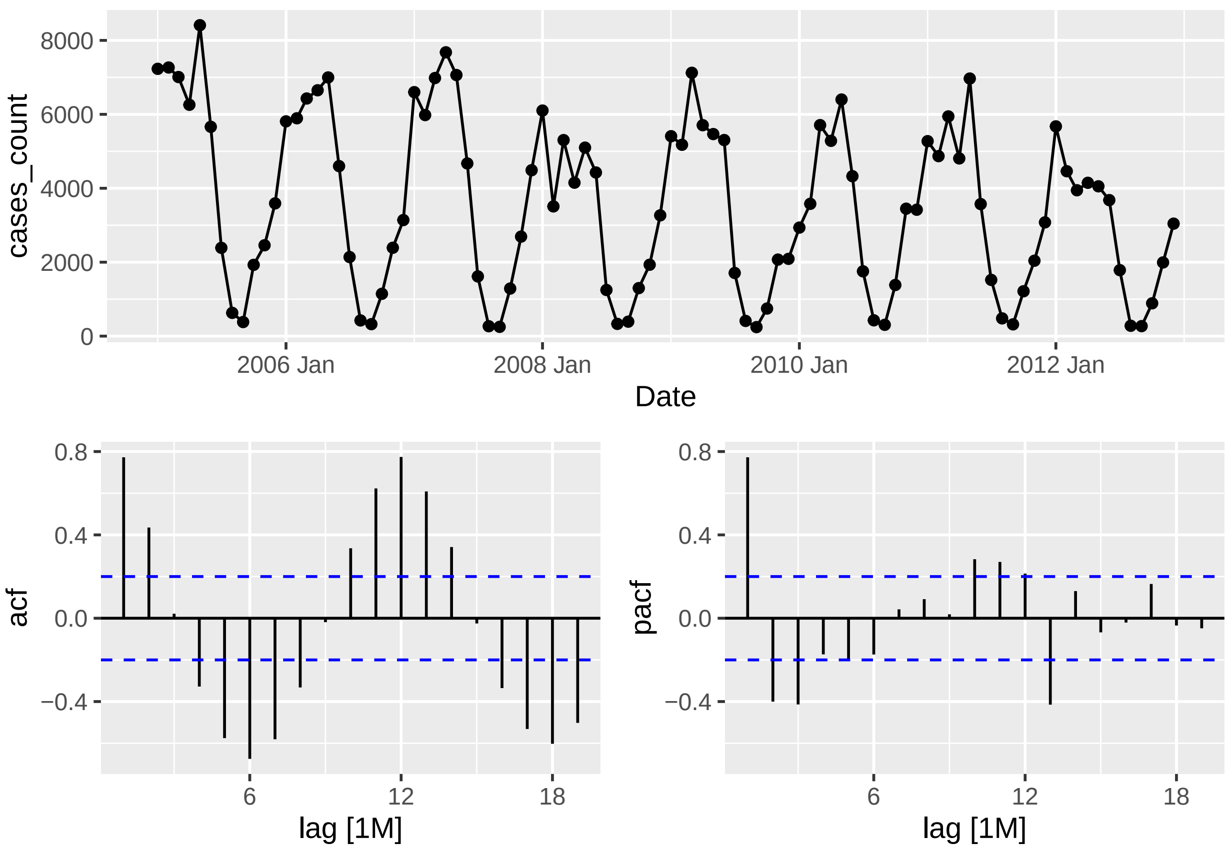
```
#Country-level training data
hch.tsb_agg_train %>%
  filter(is_aggregated(region)) %>%
  autoplot(cases_count) +
  theme(
    legend.position = 'bottom',
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)
  ) #plot the output
```



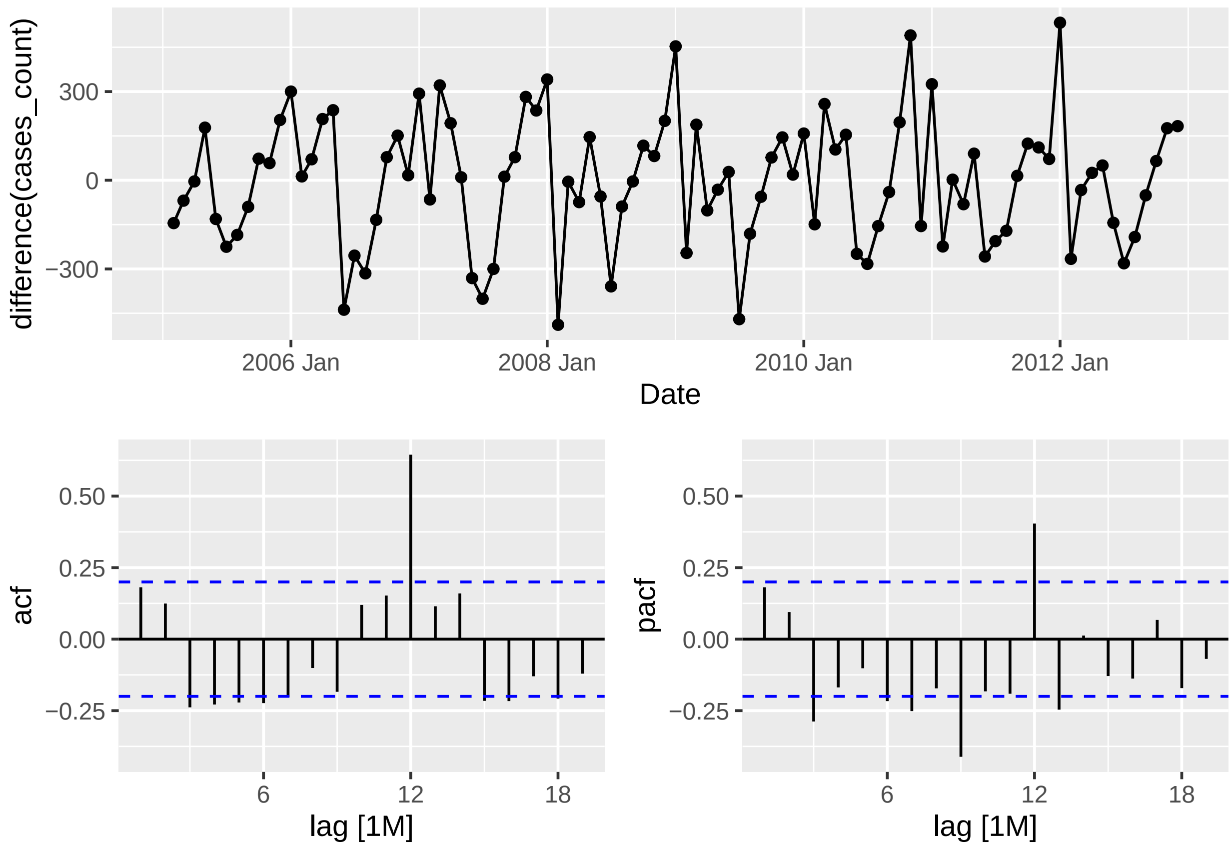

```
#Central Hungary region training data
hch.tsb_agg_train %>%
  filter(region == 'Central Hungary') %>%
  autoplot(cases_count) +
  theme(
    legend.position = 'bottom',
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)
  ) #plot the output
```




```
#Plot ACF and PACF at country and county levels.
hch.tsb_agg_train %>% filter(is_aggregated(region)) %>%  gg_tsdisplay(cases_count, plot_type='partial')
```




```
hch.tsb_agg_train %>% filter(county == 'BUDAPEST') %>%  gg_tsdisplay(difference(cases_count), plot_type=
```



4. Model ETS, ARIMA, and reconcile using bottom_up, top_down, and minimum trace methodlogies.

```
fit_all <- hch.tsb_agg_train %>%
  model(ets = ETS((cases_count)),
        arima = ARIMA((cases_count) ~ pdq(1,0,0) + PDQ(2,1,0)),
        lm = TSLM((cases_count) ~ trend() + season())) %>%
  reconcile(bu_ets = bottom_up(ets),
            td_ets = top_down(ets),
            min_trace_ets = min_trace(ets, "mint_shrink"),
            bu_arima = bottom_up(arima),
            td_arima = top_down(arima),
            min_trace_arima = min_trace(arima, "mint_shrink"),
            bu_tslm = bottom_up(lm),
            td_tslm = top_down(lm),
            min_trace_tslm = min_trace(lm, "mint_shrink")
            )
```

```
fit_all  %>%
  filter(is_aggregated(region)|county=='BUDAPEST') %>%
  select(region,county,ets,arima,lm) %>%
  pivot_longer(-c(region,county), names_to = "Model name",
                             values_to = "cases_count") %>%
  kable()
```

4.1 Base Models selected.

region	county	Model name	cases_count
Central Hungary	BUDAPEST	ets	<ETS(M,N,M)>
Central Hungary	BUDAPEST	arima	<ARIMA(1,0,0)(2,1,0)[12]>
Central Hungary	BUDAPEST	lm	
		ets	<ETS(M,N,M)>
		arima	<ARIMA(1,0,0)(2,1,0)[12] w/ drift>
		lm	

5. Analyze the IC metrics of ETS, ARIMA, and related reconcile() functions

- Country-level IC metrics:

```
fit_all %>%
  filter(is_aggregated(region))%>%
  select(region, county,ets, arima,bu_ets, bu_arima, td_ets, td_arima,lm,bu_tslm,td_tslm,min_trace_tslm
  glance() %>%
  transmute(.model,
            region = if_else(is_aggregated(region), 'country-level',as.character(region)),
            county, AICc, AIC, BIC )  %>%
  arrange(AICc) %>%
  kable()
```

.model	region	county	AICc	AIC	BIC
lm	country-level		1291.563	1286.377	1322.278
bu_tslm	country-level		1291.563	1286.377	1322.278
td_tslm	country-level		1291.563	1286.377	1322.278
min_trace_tslm	country-level		1291.563	1286.377	1322.278
arima	country-level		1370.271	1369.502	1381.656
bu_arima	country-level		1370.271	1369.502	1381.656
td_arima	country-level		1370.271	1369.502	1381.656
ets	country-level		1644.731	1638.731	1677.196
bu_ets	country-level		1644.731	1638.731	1677.196
td_ets	country-level		1644.731	1638.731	1677.196

- Budapest IC metrics:

```
fit_all %>%
  filter(county=='BUDAPEST')%>%
  select(region, county,ets, arima,bu_ets, bu_arima, td_ets,
         td_arima,lm,bu_tslm,td_tslm,min_trace_tslm ) %>%
  glance() %>%
  transmute(.model,
            region = if_else(is_aggregated(region), 'country-level',as.character(region)),
            county, AICc, AIC, BIC )  %>%
  arrange(AICc) %>%
  kable()
```

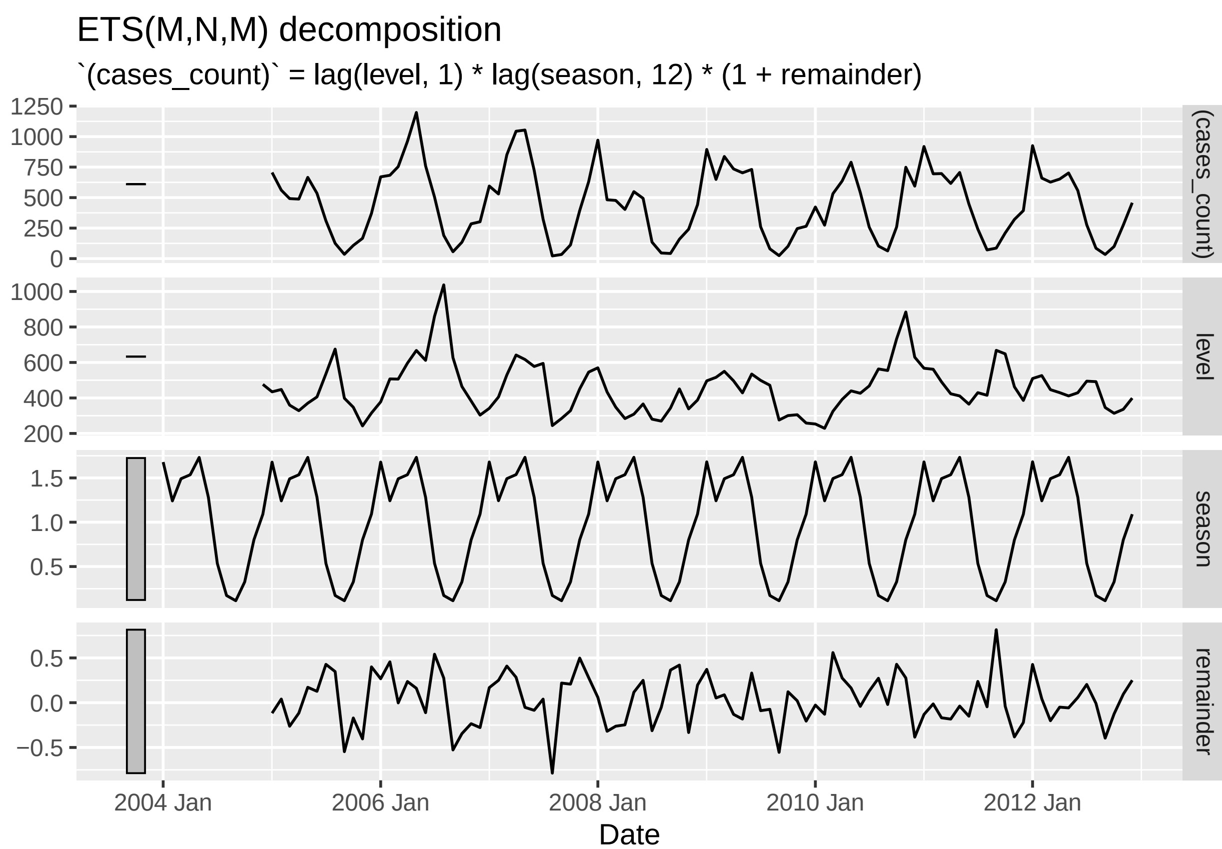
.model	region	county	AICc	AIC	BIC
lm	Central Hungary	BUDAPEST	977.3666	972.1814	1008.082
bu_tslm	Central Hungary	BUDAPEST	977.3666	972.1814	1008.082

.model	region	county	AICc	AIC	BIC
td_tslm	Central Hungary	BUDAPEST	977.3666	972.1814	1008.082
min_trace_tslm	Central Hungary	BUDAPEST	977.3666	972.1814	1008.082
arima	Central Hungary	BUDAPEST	1052.6526	1052.1463	1061.870
bu_arima	Central Hungary	BUDAPEST	1052.6526	1052.1463	1061.870
td_arima	Central Hungary	BUDAPEST	1052.6526	1052.1463	1061.870
ets	Central Hungary	BUDAPEST	1344.9748	1338.9748	1377.440
bu_ets	Central Hungary	BUDAPEST	1344.9748	1338.9748	1377.440
td_ets	Central Hungary	BUDAPEST	1344.9748	1338.9748	1377.440

6. Get the decomposition of the ets

- Budapest (county level) decomposition:

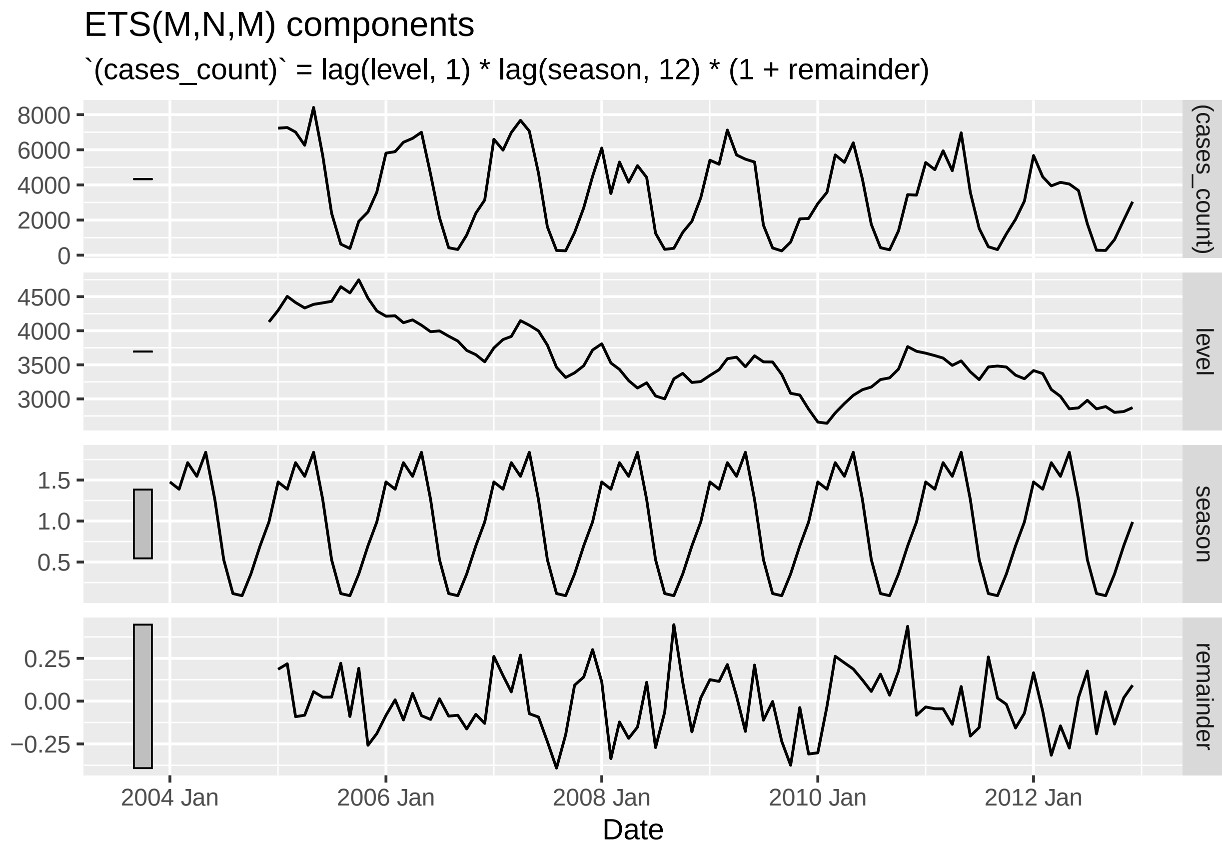
```
fit_all %>%
  filter(county=='BUDAPEST') %>%
  select(td_ets) %>%
  components() %>%
  autoplot()
```



- Country-level decomposition:

```
fit_all %>%
  filter(is_aggregated(region)) %>%
  select(ets) %>%
  components() %>%
```

```
  autoplot() +
  labs(title = "ETS(M,N,M) components")
```



7. Run the forecast of next 2 years on the model:

```
fc_all <- fit_all %>%
  forecast(h=24)
```

7.1 Get the accuracy metrics for comparison:

- Country-level accuracy metrics:

```
fc_all %>%
  accuracy(
    data = hch.tsb_agg,
    measures = list(rmse = RMSE, mase = MASE, mape = MAPE, mae=MAE)
  ) %>%
  filter(is_aggregated(region)) %>%
  arrange(rmse) %>%
  transmute(.model,
            region = if_else(is_aggregated(region),
                             'country-level',
                             as.character(region)),
            county, rmse, mase, mape, mae) %>%
  kable()
```

.model	region	county	rmse	mase	mape	mae
ets	country-level		632.3692	0.6092038	21.35646	451.4708
td_ets	country-level		632.3692	0.6092038	21.35646	451.4708
min_trace_ets	country-level		641.9707	0.6257824	21.92544	463.7569
bu_ets	country-level		651.9269	0.6409759	22.25146	475.0166
arima	country-level		667.4608	0.7509349	41.25541	556.5053
td_arima	country-level		667.4608	0.7509349	41.25541	556.5053
min_trace_arima	country-level		792.5440	0.7547412	27.81123	559.3261
bu_arima	country-level		836.7337	0.7924584	27.18954	587.2777
bu_tslm	country-level		851.6466	0.9436430	63.88471	699.3181
lm	country-level		851.6466	0.9436430	63.88471	699.3181
td_tslm	country-level		851.6466	0.9436430	63.88471	699.3181
min_trace_tslm	country-level		851.6466	0.9436430	63.88471	699.3181

- Budapest (county-level) accuracy metrics:

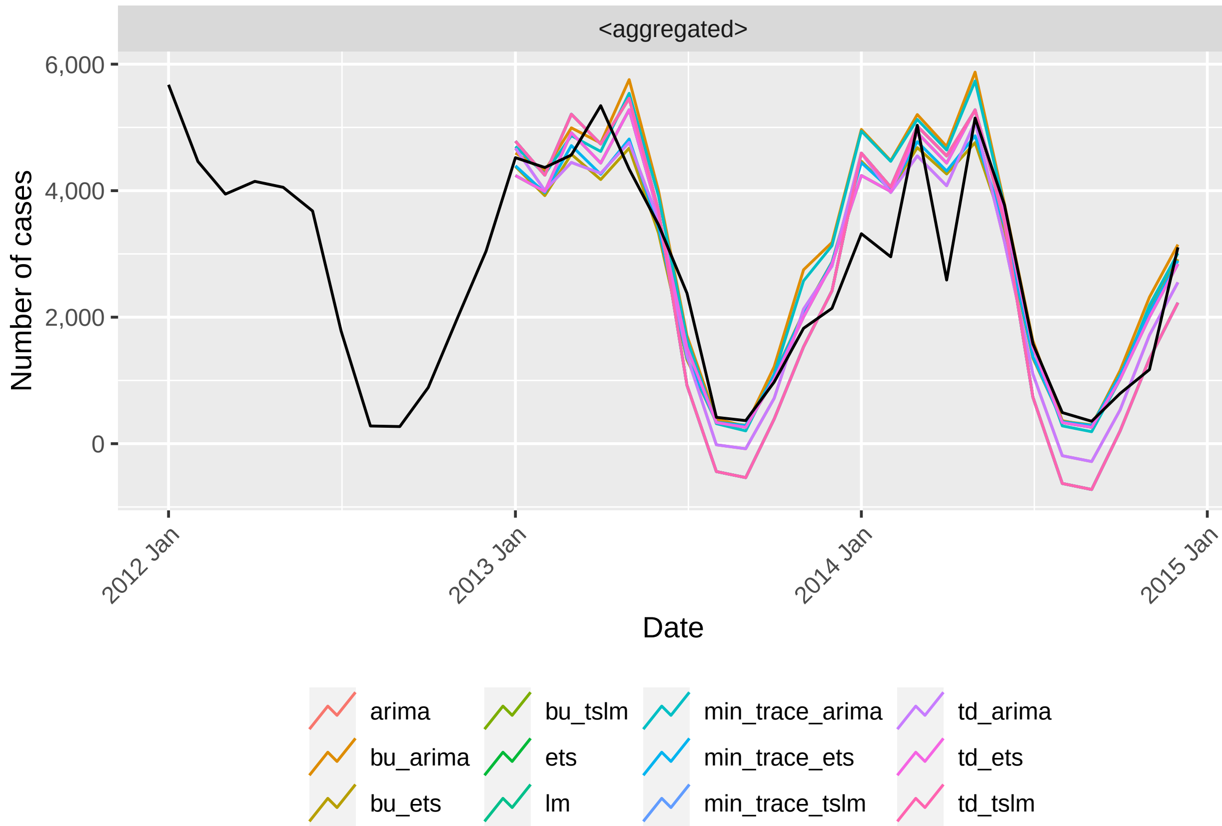
```
fc_all %>%
  accuracy(
    data = hch.tsb_agg,
    measures = list(rmse = RMSE, mase = MASE, mape = MAPE, mae=MAE)
  ) %>%
  filter(county=='BUDAPEST') %>%
  arrange(rmse) %>%
  transmute(.model,
            region = if_else(is_aggregated(region),
                             'country-level',
                             as.character(region)),
            county, rmse, mase, mape, mae) %>%
  kable()
```

.model	region	county	rmse	mase	mape	mae
td_ets	Central Hungary	BUDAPEST	136.3707	0.6644929	33.75123	105.0927
min_trace_ets	Central Hungary	BUDAPEST	141.3894	0.6747999	34.33903	106.7228
bu_ets	Central Hungary	BUDAPEST	142.6083	0.6789840	34.88513	107.3846
ets	Central Hungary	BUDAPEST	142.6083	0.6789840	34.88513	107.3846
min_trace_tslm	Central Hungary	BUDAPEST	147.0184	0.6699954	38.10158	105.9630
bu_tslm	Central Hungary	BUDAPEST	147.0184	0.6699954	38.10158	105.9630
lm	Central Hungary	BUDAPEST	147.0184	0.6699954	38.10158	105.9630
td_tslm	Central Hungary	BUDAPEST	147.0184	0.6699954	38.10158	105.9630
td_arima	Central Hungary	BUDAPEST	171.4101	0.8742857	61.82152	138.2725
min_trace_arima	Central Hungary	BUDAPEST	185.8736	0.8356473	50.90250	132.1616
arima	Central Hungary	BUDAPEST	186.2432	0.8359951	50.80321	132.2166
bu_arima	Central Hungary	BUDAPEST	186.2432	0.8359951	50.80321	132.2166

8. Plot the forecast at the country-level:

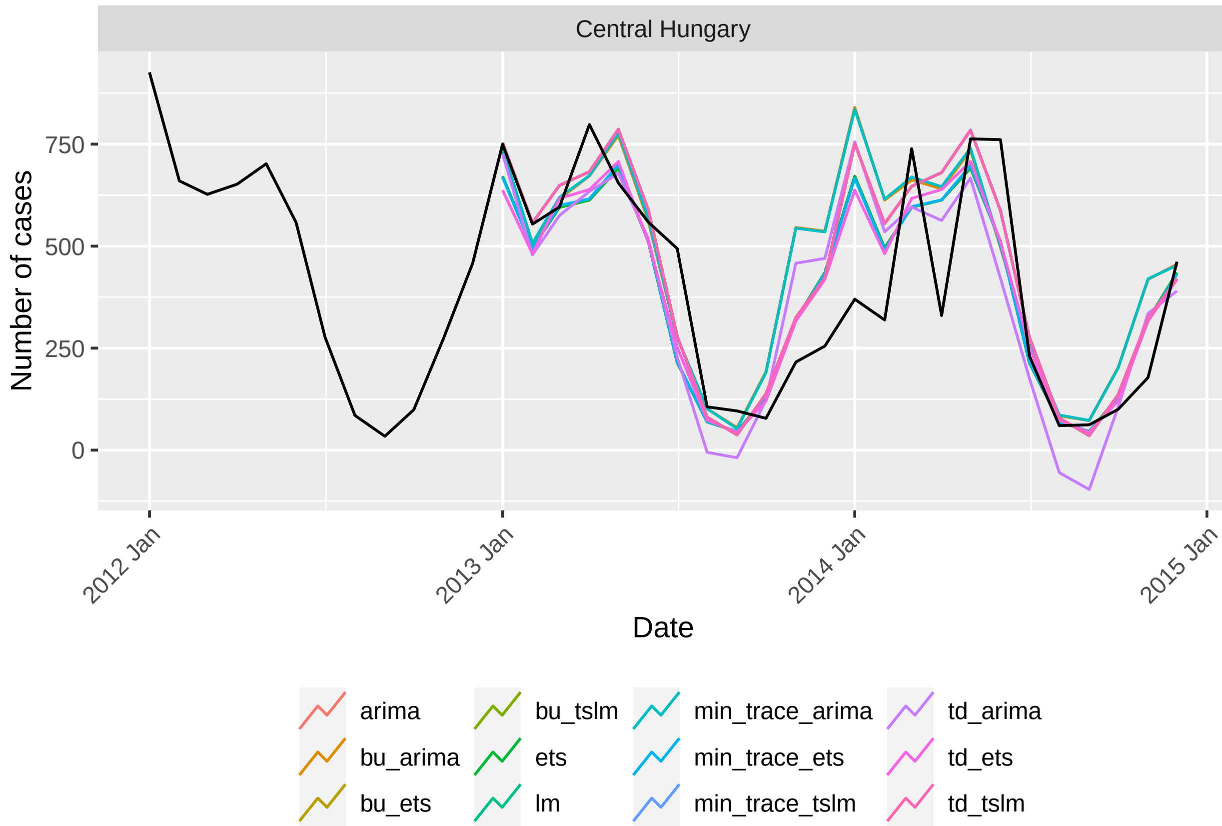
```
autoplot(
  fc_all %>%
    filter( is_aggregated(region)),
  hch.tsb_agg %>%
    ungroup() %>%
    filter(year(Date) >2011), level = NULL) +
  facet_wrap(~region, scales = "free_y") +
```

```
  scale_y_continuous(labels = scales::comma_format()) +
  labs(color = "", x = "Date" ,y = "Number of cases") +
  theme(
    legend.position = 'bottom',
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)
  )
```



Plot the forecast at the country-level and all the regions

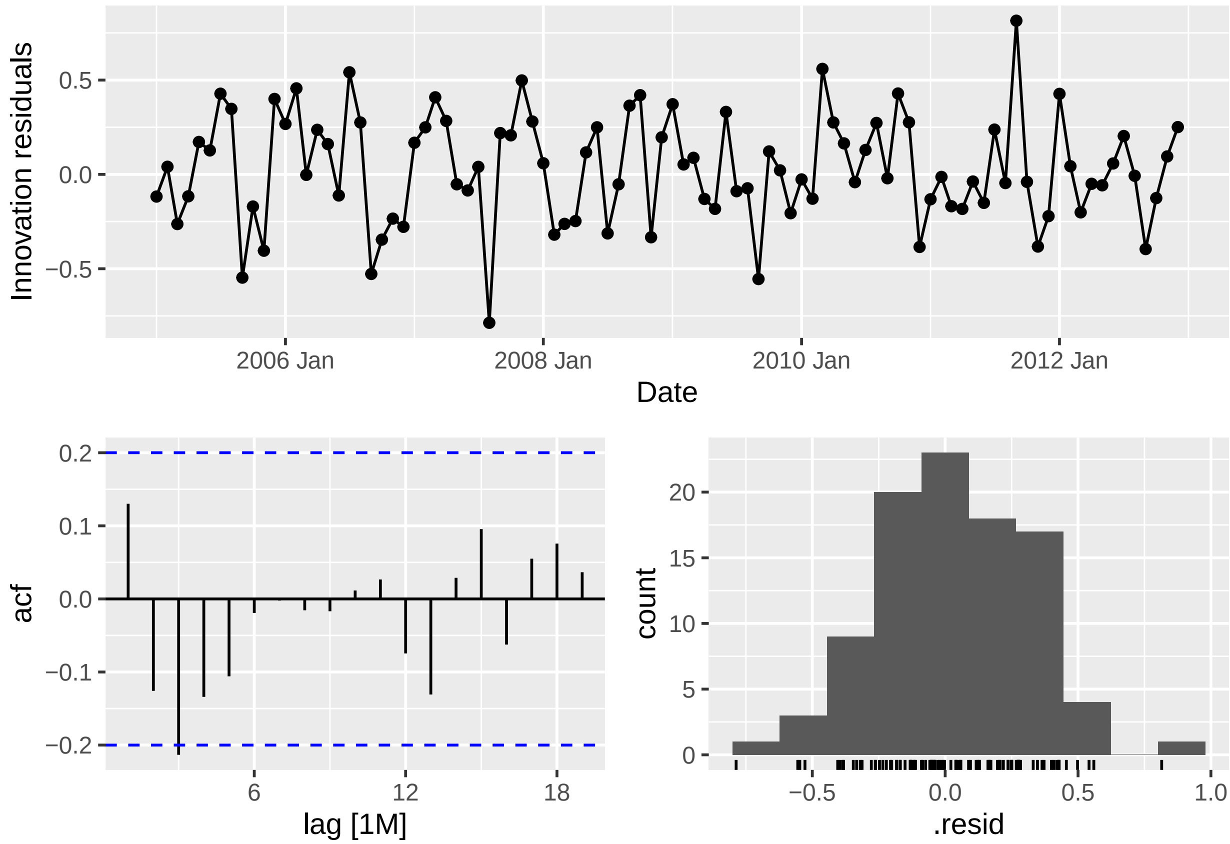
```
autoplot(
  fc_all %>%
    filter(county=='BUDAPEST' ),
  hch.tsb_agg %>%
    ungroup() %>%
    filter(year(Date) >2011), level = NULL) +
  facet_wrap(~region, scales = "free_y") +
  scale_y_continuous(labels = scales::comma_format()) +
  labs(color = "", x = "Date" ,y = "Number of cases") +
  theme(
    legend.position = 'bottom',
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)
  )
```



7.2 Get the ACF plot of the residuals of top two models based on RMSE at county level:

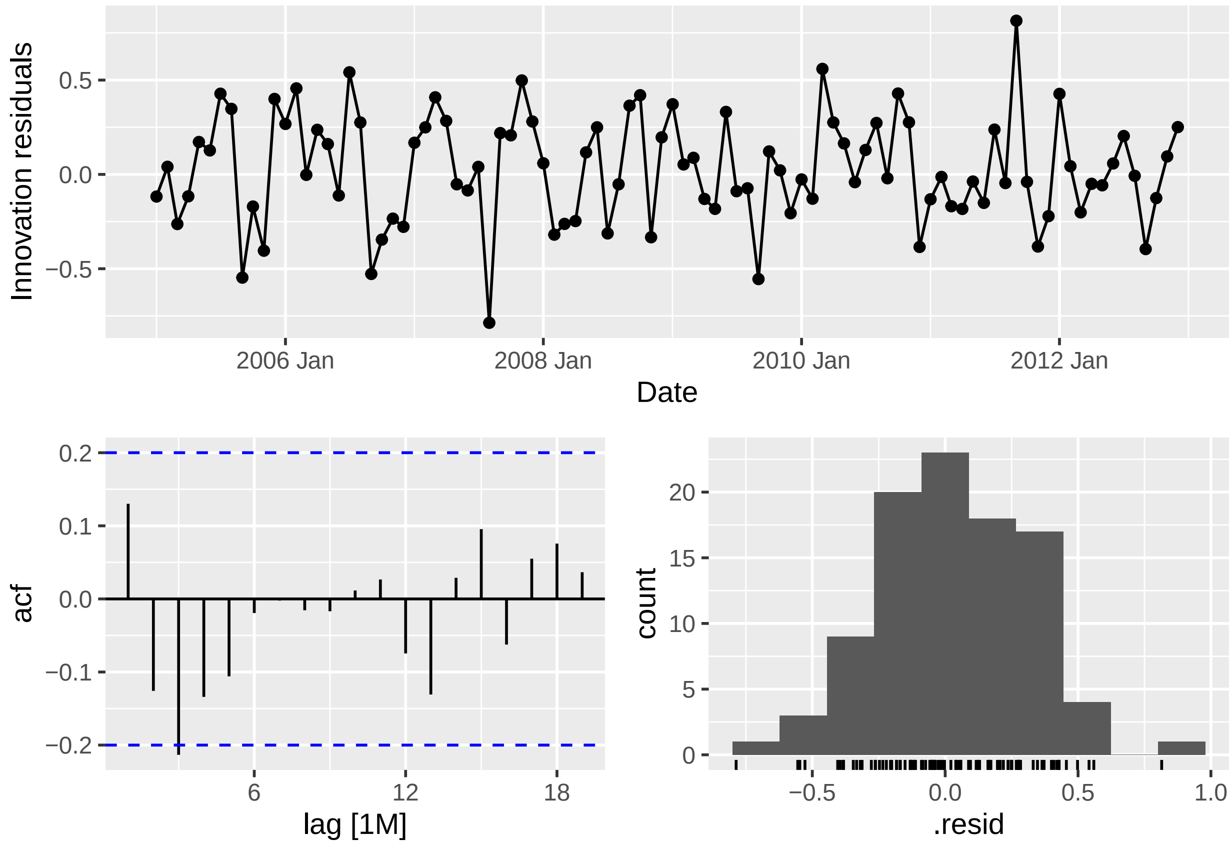
- ETS based on Top-down approach:

```
fit_all  %>%
   filter(county=='BUDAPEST' ) %>%
  select(td_ets)  %>%
  gg_tsresiduals()
```



- Base ETS plot:

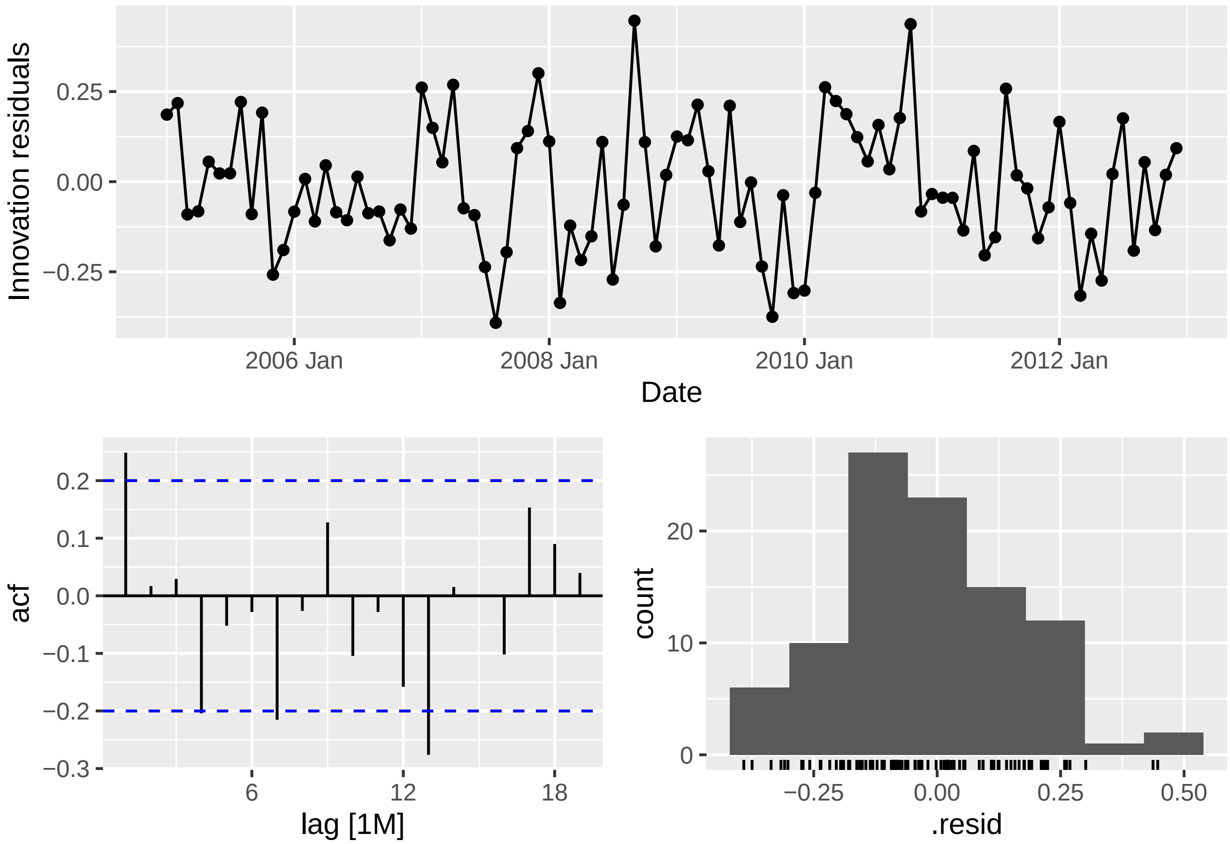
```
fit_all  %>%
   filter(county=='BUDAPEST' ) %>%
  select(min_trace_ets)  %>%
  gg_tsresiduals()
```



7.3 Get the ACF plot of the residuals of top two models based on RMSE at COUNTRY level:

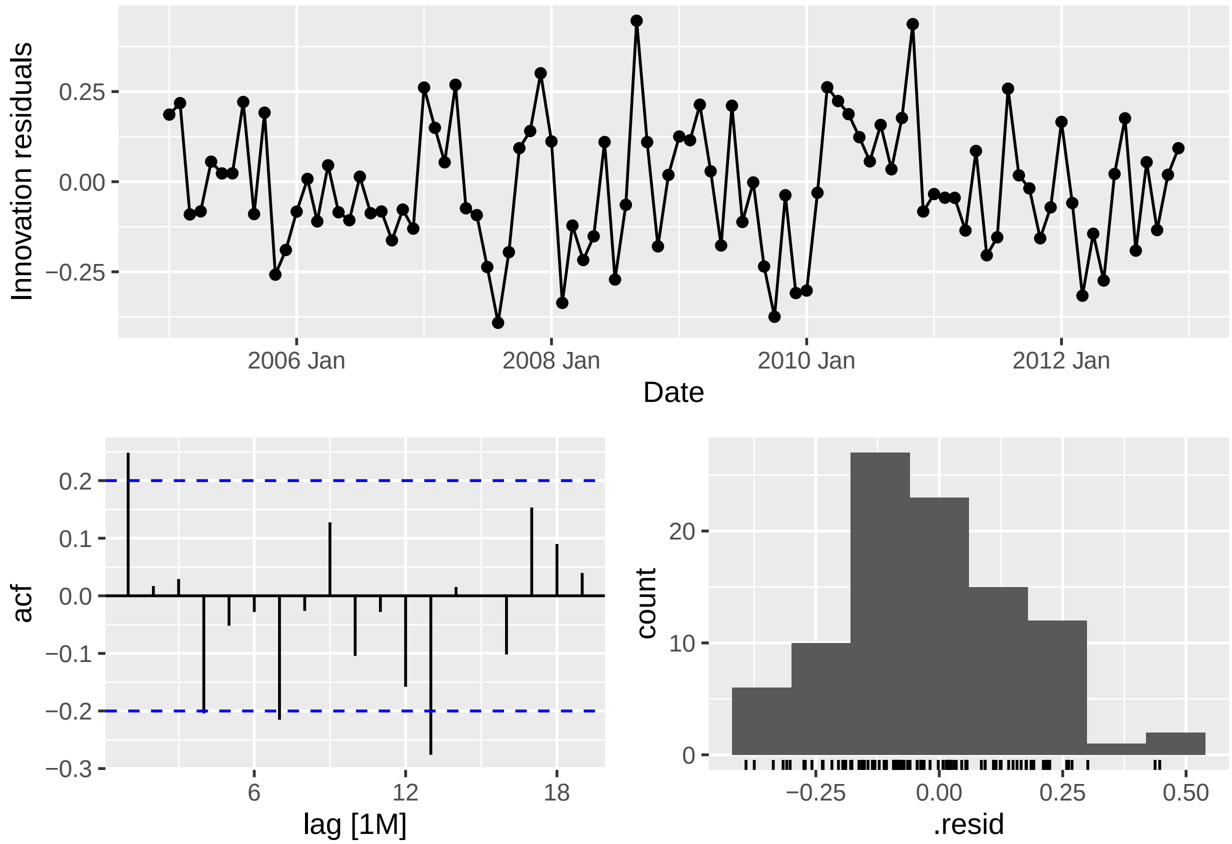
- Base ETS:

```
fit_all  %>%
   filter(is_aggregated(region) ) %>%
  select(ets)  %>%
  gg_tsresiduals()
```



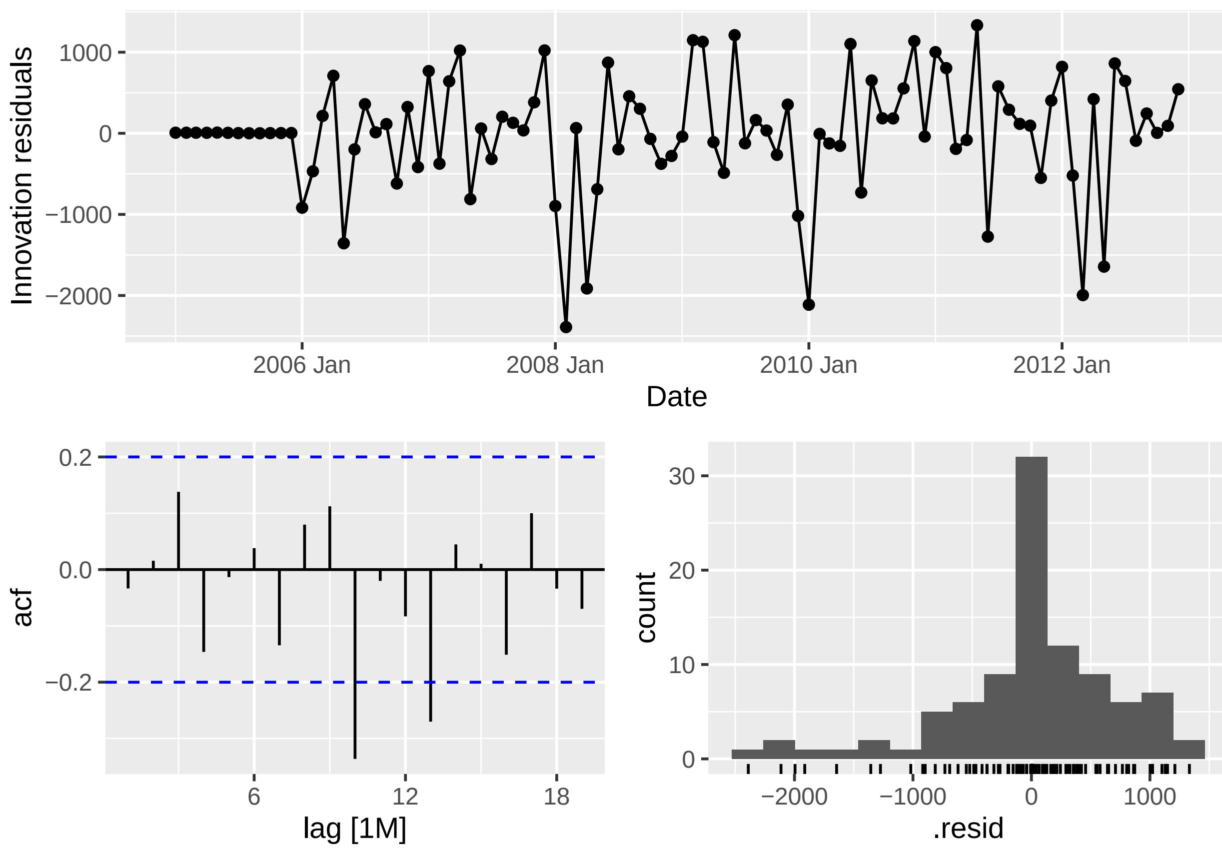
- ETS based on Top-down approach:

```
fit_all  %>%
   filter(is_aggregated(region) ) %>%
  select(td_ets)  %>%
  gg_tsresiduals()
```



- Base ARIMA:

```
fit_all  %>%
   filter(is_aggregated(region) ) %>%
  select(arima)  %>%
  gg_tsresiduals()
```



7.4 Conduct Ljung-Box test on the fit:

- Ljung-Box test: we see a p-value much smaller than 0.05, thus we can reject the null hypothesis, indicating the time series does contain an autocorrelation.

```
augment(fit_all)  %>%
  filter(is_aggregated(region) | county=='BUDAPEST') %>%
  filter(.model=='ets'|.model=='arima'|.model=='td_ets'|.model=='min_trace_ets')  %>%
  features(.resid, ljung_box,lag=12) %>%
  kable()
```

region	county	.model	lb_stat	lb_pvalue
Central Hungary	BUDAPEST	arima	8.66519	0.7312303
Central Hungary	BUDAPEST	ets	11.36668	0.4977826
Central Hungary	BUDAPEST	min_trace_ets	11.36668	0.4977826
Central Hungary	BUDAPEST	td_ets	11.36668	0.4977826
		arima	21.55678	0.0427981
		ets	22.16698	0.0356904
		min_trace_ets	22.16698	0.0356904
		td_ets	22.16698	0.0356904