Mort against Mort Change Comparer

# Aim

The aim of this is to compare mortality against change in mortality for a number of countries.

## Load the packages

pacman::p\_load(  
 tidyverse,   
 plotly,  
 HMDHFDplus  
)

Now, how do I grab life expectancies only for each country?

# username <- "jon.will.minton@gmail.com"  
# password <- "passphrase"  
#   
# hmd\_dta <- data\_frame(  
# code = getHMDcountries()  
# ) %>%   
# mutate(deaths = map(code, readHMDweb, item = "e0per", username = username, password = password))

API approach seems to be blocked in Meridian Court (Firewall?) so have manually downloaded

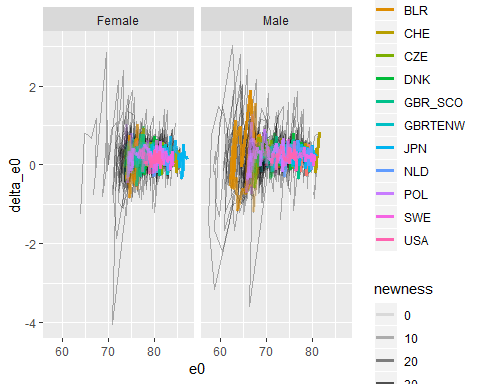
baseloc <- "N:/Jon/hmd\_data/hmd\_countries/"  
  
codes <- dir(baseloc)  
  
hmd\_e0 <- data\_frame(  
 code = codes  
) %>%   
 mutate(  
 e0 = map(  
 code,  
 function(x) {  
 read\_table(  
 file.path(  
 baseloc, x, "STATS/e0per.txt"  
 ),  
 skip = 2,  
 col\_types = "iddd"  
   
 )  
 }  
 )  
 ) %>%   
 unnest() %>%   
 gather(key = "gender", value = "e0", Female:Total)  
  
  
# Code to make this data available for app  
# names(hmd\_e0) <- tolower(names(hmd\_e0))  
#   
# hmd\_e0 <- hmd\_e0 %>%   
# mutate(gender = tolower(gender))  
#   
# write\_csv(x = hmd\_e0, path = "hmd\_explorer/data/hmd\_e0.csv")

Now to visualise

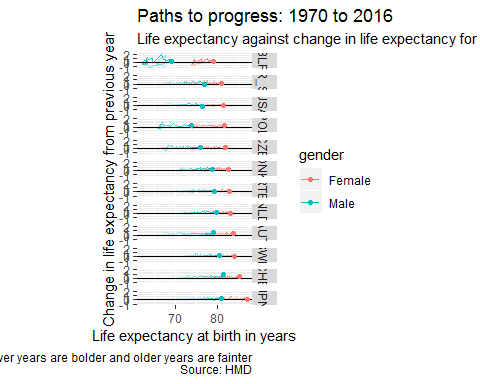
hmd\_e0 %>%   
 group\_by(code, gender) %>%   
 arrange(Year) %>%   
 mutate(delta\_e0 = e0 - lag(e0)) %>%   
 filter(!is.na(delta\_e0)) %>%   
 filter(gender != "Total") %>%   
 filter(Year >= 1990) -> tmp  
  
tmp %>%   
 ggplot(aes(y = delta\_e0, x = e0, group = code)) +   
 geom\_path(alpha = 0.3) +  
 facet\_wrap(~gender) +  
 geom\_path(  
 aes(y = delta\_e0, x = e0, colour = code, alpha = newness),  
 size = 1.2,  
 data = tmp %>%   
 filter(code %in% c("GBR\_SCO", "USA", "GBRTENW")) %>%   
 mutate(newness = Year - 1990),  
 inherit.aes = FALSE  
 )



hmd\_e0 %>%   
 group\_by(code, gender) %>%   
 arrange(Year) %>%   
 mutate(delta\_e0 = e0 - lag(e0)) %>%   
 filter(!is.na(delta\_e0)) %>%   
 filter(gender != "Total") %>%   
 filter(Year >= 1970) -> tmp  
  
tmp %>%   
 ggplot(aes(y = delta\_e0, x = e0, group = code)) +   
 geom\_path(alpha = 0.3) +  
 facet\_wrap(~gender) +  
 geom\_path(  
 aes(y = delta\_e0, x = e0, colour = code, alpha = newness),  
 size = 1.2,  
 data = tmp %>%   
 filter(code %in% c("GBR\_SCO", "USA", "GBRTENW", "NLD", "FRANTP", "SWE", "AUT", "CHE", "POL", "DNK", "CZE", "JPN", "BLR")) %>%   
 mutate(newness = Year - 1970),  
 inherit.aes = FALSE  
 )



tmp %>%   
 filter(code %in% c("GBR\_SCO", "USA", "GBRTENW", "NLD", "FRANTP", "SWE", "AUT", "CHE", "POL", "DNK", "CZE", "JPN", "BLR")) %>%   
 group\_by(code) %>%   
 mutate(val\_2016 = e0[Year == 2016 & gender == "Female"]) %>%   
 ungroup() %>%   
 mutate(newness = Year - 1970) %>%   
 mutate(code = fct\_reorder(code, val\_2016)) %>%   
 ggplot(aes(y = delta\_e0, x = e0, alpha = newness, colour = gender)) +   
 geom\_path() +  
 facet\_grid(code~.) +  
 geom\_hline(yintercept = 0) +  
 coord\_equal() +  
 geom\_point(  
 aes(y = delta\_e0, x = e0, colour = gender),  
 inherit.aes = F,  
 data =   
 tmp %>%   
 filter(code %in% c("GBR\_SCO", "USA", "GBRTENW", "NLD", "FRANTP", "SWE", "AUT", "CHE", "POL", "DNK", "CZE", "JPN", "BLR")) %>%   
 group\_by(code) %>%   
 mutate(val\_2016 = e0[Year == 2016 & gender == "Female"]) %>%   
 ungroup() %>%   
 mutate(code = fct\_reorder(code, val\_2016)) %>%   
 filter(Year == 2016)  
 ) +  
 labs(x = "Life expectancy at birth in years", y = "Change in life expectancy from previous year",  
 title = "Paths to progress: 1970 to 2016",  
 subtitle = "Life expectancy against change in life expectancy for selected countries (Dot indicates 2016)",  
 caption = "Newer years are bolder and older years are fainter\nSource: HMD") +   
 guides(alpha = FALSE) +  
 theme(  
 plot.caption = element\_text(hjust = 1)  
 )

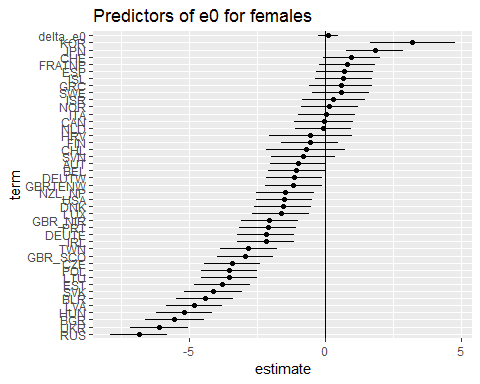


ggsave("sc.png", height = 30, width = 20, units = "cm", dpi = 300)

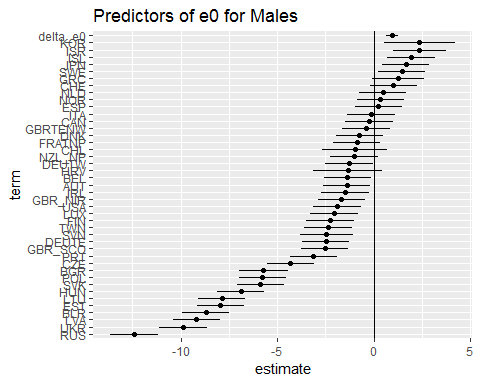
# Modelling

After having shown the intuition, the next stage is to model the relationship between delta(e0) and e0 for all available countries

tmp %>%   
 filter(gender == "Female") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0 + code, data = .) %>%   
 summary() %>%   
 broom::tidy() %>%   
 filter(term != "(Intercept)") %>%   
 mutate(term = str\_replace(term, "code", "")) %>%   
 mutate(term = fct\_reorder(term, estimate)) %>%   
 mutate(term = fct\_relevel(term, "delta\_e0", after = Inf)) %>%   
 ggplot() +   
 geom\_point(aes(x = estimate, y= term)) +   
 geom\_segment(aes(x = estimate - 1.96 \* std.error, xend = estimate + 1.96 \* std.error, y = term, yend = term)) +   
 geom\_vline(xintercept = 0) +   
 labs(title = "Predictors of e0 for females")



tmp %>%   
 filter(gender == "Male") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0 + code, data = .) %>%   
 summary() %>%   
 broom::tidy() %>%   
 filter(term != "(Intercept)") %>%   
 mutate(term = str\_replace(term, "code", "")) %>%   
 mutate(term = fct\_reorder(term, estimate)) %>%   
 mutate(term = fct\_relevel(term, "delta\_e0", after = Inf)) %>%   
 ggplot() +   
 geom\_point(aes(x = estimate, y= term)) +   
 geom\_segment(aes(x = estimate - 1.96 \* std.error, xend = estimate + 1.96 \* std.error, y = term, yend = term)) +   
 geom\_vline(xintercept = 0) +   
 labs(title = "Predictors of e0 for Males")

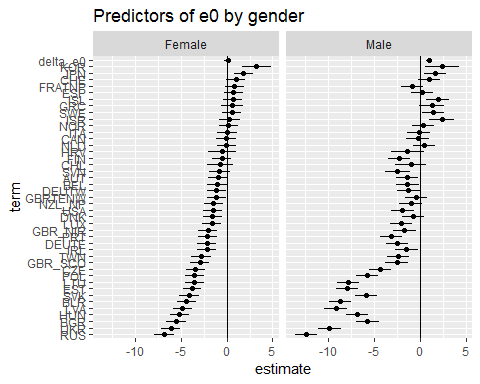


And now for both, given the range is much greater for males than females.

results\_females <- tmp %>%   
 filter(gender == "Female") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0 + code, data = .) %>%   
 summary() %>%   
 broom::tidy() %>%   
 filter(term != "(Intercept)") %>%   
 mutate(term = str\_replace(term, "code", "")) %>%   
 mutate(sex = "Female") %>%  
 mutate(term = fct\_reorder(term, estimate)) %>%   
 mutate(term = fct\_relevel(term, "delta\_e0", after = Inf))  
  
results\_males <- tmp %>%   
 filter(gender == "Male") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0 + code, data = .) %>%   
 summary() %>%   
 broom::tidy() %>%   
 filter(term != "(Intercept)") %>%   
 mutate(term = str\_replace(term, "code", "")) %>%   
 mutate(sex = "Male")  
  
results\_males$term <- factor(results\_males$term, levels = levels(results\_females$term))  
  
results\_both <- results\_females %>%   
 bind\_rows(results\_males)

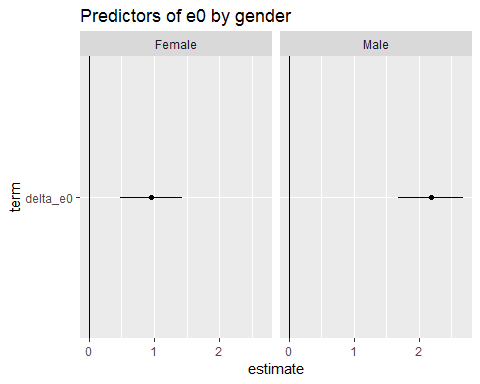
And now to visualise

results\_both %>%   
 ggplot(aes(group = sex)) +   
 geom\_point(aes(x = estimate, y= term)) +   
 geom\_segment(aes(x = estimate - 1.96 \* std.error, xend = estimate + 1.96 \* std.error, y = term, yend = term)) +   
 geom\_vline(xintercept = 0) +   
 labs(title = "Predictors of e0 by gender") +   
 facet\_wrap(~sex)



However, what’s presented above is the result of a ‘level 2’ model. The simplest, ‘level 1’ model, would be one that simply regresses e0 on delta\_e0 and nothing else. What does this show?

results\_females <- tmp %>%   
 filter(gender == "Female") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0, data = .) %>%   
 summary() %>%   
 broom::tidy() %>%   
 filter(term != "(Intercept)") %>%   
 mutate(term = str\_replace(term, "code", "")) %>%   
 mutate(sex = "Female") %>%  
 mutate(term = fct\_reorder(term, estimate)) %>%   
 mutate(term = fct\_relevel(term, "delta\_e0", after = Inf))  
  
results\_males <- tmp %>%   
 filter(gender == "Male") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0, data = .) %>%   
 summary() %>%   
 broom::tidy() %>%   
 filter(term != "(Intercept)") %>%   
 mutate(term = str\_replace(term, "code", "")) %>%   
 mutate(sex = "Male")  
  
results\_males$term <- factor(results\_males$term, levels = levels(results\_females$term))  
  
results\_both <- results\_females %>%   
 bind\_rows(results\_males)  
  
results\_both %>%   
 ggplot(aes(group = sex)) +   
 geom\_point(aes(x = estimate, y= term)) +   
 geom\_segment(aes(x = estimate - 1.96 \* std.error, xend = estimate + 1.96 \* std.error, y = term, yend = term)) +   
 geom\_vline(xintercept = 0) +   
 labs(title = "Predictors of e0 by gender") +   
 facet\_wrap(~sex)



So, this changes the coefficient for delta\_e0, for females, from not statistically significant, to statistically significant, but perhaps what’s more important to establish is whether the country codes should be included in terms of improvement to model fit. This question can be answered using ANOVA, as the second model is nested in the first model.

i.e. the model without the country fixed effects is the *unrestricted model*, and the model with the country fixed effects is the *restricted model*.

model\_restricted\_f <- tmp %>%   
 filter(gender == "Female") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0, data = .)  
  
model\_restricted\_m <- tmp %>%   
 filter(gender == "Male") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0, data = .)  
  
model\_unrestricted\_f <- tmp %>%   
 filter(gender == "Female") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0 + code, data = .)  
  
model\_unrestricted\_m <- tmp %>%   
 filter(gender == "Male") %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 lm(e0 ~ delta\_e0 + code, data = .)  
  
# Test for females  
anova(model\_restricted\_f, model\_unrestricted\_f)

## Analysis of Variance Table  
##   
## Model 1: e0 ~ delta\_e0  
## Model 2: e0 ~ delta\_e0 + code  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 1840 20072   
## 2 1798 11613 42 8458.1 31.178 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Test for males  
anova(model\_restricted\_m, model\_unrestricted\_m)

## Analysis of Variance Table  
##   
## Model 1: e0 ~ delta\_e0  
## Model 2: e0 ~ delta\_e0 + code  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 1840 39024   
## 2 1798 16053 42 22971 61.26 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

As expected, the ANOVA tests suggest the unrestricted model variants are justified by the better model fit for both genders (especially males).

We can extend the approach of model comparison to see if adding a post 2010 dummy term is justified

tmp %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 mutate(is\_after\_2010 = Year >= 2010) %>%   
 group\_by(gender) %>%   
 nest() %>%   
 mutate(  
 mod\_01 = map(data, ~lm(e0 ~ delta\_e0, data = .x)),  
 mod\_02 = map(data, ~lm(e0 ~ delta\_e0 + code, data = .x)),  
 mod\_03 = map(data, ~lm(e0 ~ delta\_e0 + code + is\_after\_2010, data = .x)),  
 mod\_04 = map(data, ~lm(e0 ~ delta\_e0 + is\_after\_2010, data = .x))  
 ) %>%   
 select(-data) %>%   
 gather(key = "mod\_label", value = "model", -gender) %>%  
 mutate(  
 aic = map\_dbl(model, AIC),   
 bic = map\_dbl(model, BIC)  
 ) %>%   
 group\_by(gender) %>%   
 mutate(fit\_rank\_a = rank(aic)) %>%   
 mutate(fit\_rank\_b = rank(bic)) %>%   
 ungroup() %>%   
 arrange(gender, fit\_rank\_a)

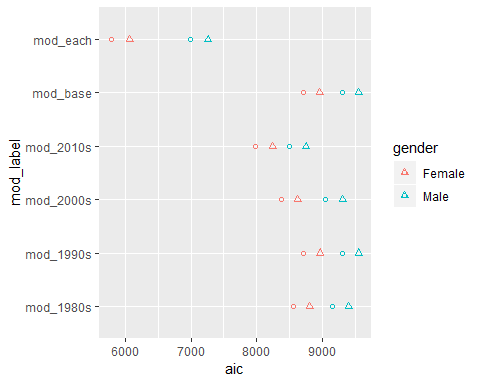
## # A tibble: 8 x 7  
## gender mod\_label model aic bic fit\_rank\_a fit\_rank\_b  
## <chr> <chr> <list> <dbl> <dbl> <dbl> <dbl>  
## 1 Female mod\_03 <S3: lm> 7988. 8242. 1 1  
## 2 Female mod\_02 <S3: lm> 8709. 8957. 2 2  
## 3 Female mod\_04 <S3: lm> 9217. 9240. 3 3  
## 4 Female mod\_01 <S3: lm> 9633. 9649. 4 4  
## 5 Male mod\_03 <S3: lm> 8500. 8754. 1 1  
## 6 Male mod\_02 <S3: lm> 9305. 9554. 2 2  
## 7 Male mod\_04 <S3: lm> 10571. 10593. 3 3  
## 8 Male mod\_01 <S3: lm> 10858. 10874. 4 4

Let’s now extend this kind of approach to other decadal periods

timeblock\_models <- tmp %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%  
 mutate(is\_1980s = Year >= 1980 & Year <= 1989) %>%   
 mutate(is\_1990s = Year >= 1990 & Year <= 1999) %>%   
 mutate(is\_2000s = Year >= 2000 & Year <= 2009) %>%   
 mutate(is\_after\_2010 = Year >= 2010) %>%   
 group\_by(gender) %>%   
 nest() %>%   
 mutate(  
 mod\_base = map(data, ~lm(e0 ~ delta\_e0 + code, data = .x)),  
 mod\_2010s = map(data, ~lm(e0 ~ delta\_e0 + code + is\_after\_2010, data = .x)),  
 mod\_2000s = map(data, ~lm(e0 ~ delta\_e0 + code + is\_2000s, data = .x)),  
 mod\_1990s = map(data, ~lm(e0 ~ delta\_e0 + code + is\_1990s, data = .x)),  
 mod\_1980s = map(data, ~lm(e0 ~ delta\_e0 + code + is\_1980s, data = .x)),  
 mod\_each = map(data, ~lm(e0 ~ delta\_e0 + code + is\_1980s + is\_1990s + is\_2000s + is\_after\_2010, data = .x))  
 ) %>%   
 select(-data) %>%   
 gather(key = "mod\_label", value = "model", -gender) %>%  
 mutate(  
 aic = map\_dbl(model, AIC),   
 bic = map\_dbl(model, BIC)  
 ) %>%   
 group\_by(gender) %>%   
 mutate(fit\_rank\_a = rank(aic)) %>%   
 mutate(fit\_rank\_b = rank(bic)) %>%   
 ungroup() %>%   
 arrange(gender, fit\_rank\_a)   
  
timeblock\_models

## # A tibble: 12 x 7  
## gender mod\_label model aic bic fit\_rank\_a fit\_rank\_b  
## <chr> <chr> <list> <dbl> <dbl> <dbl> <dbl>  
## 1 Female mod\_each <S3: lm> 5795. 6066. 1 1  
## 2 Female mod\_2010s <S3: lm> 7988. 8242. 2 2  
## 3 Female mod\_2000s <S3: lm> 8375. 8628. 3 3  
## 4 Female mod\_1980s <S3: lm> 8553. 8807. 4 4  
## 5 Female mod\_base <S3: lm> 8709. 8957. 5 5  
## 6 Female mod\_1990s <S3: lm> 8711. 8965. 6 6  
## 7 Male mod\_each <S3: lm> 6994. 7264. 1 1  
## 8 Male mod\_2010s <S3: lm> 8500. 8754. 2 2  
## 9 Male mod\_2000s <S3: lm> 9053. 9307. 3 3  
## 10 Male mod\_1980s <S3: lm> 9150. 9404. 4 4  
## 11 Male mod\_1990s <S3: lm> 9300. 9554. 5 6  
## 12 Male mod\_base <S3: lm> 9305. 9554. 6 5

timeblock\_models %>%   
 ggplot(aes(y = mod\_label, colour = gender, group = gender)) +   
 geom\_point(aes(x = aic), shape = 1) + geom\_point(aes(x = bic), shape = 2)

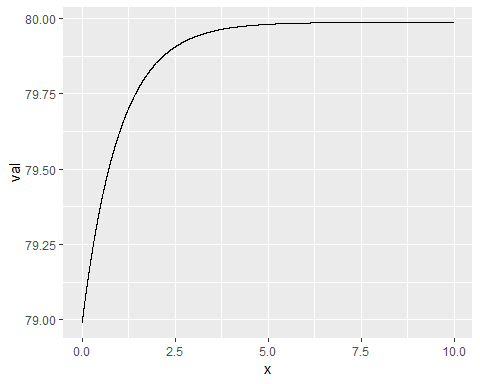


Generally, the models with newer dummy terms outperformed those with older dummy terms, though for males the 1980s dummy model outperformed the 1990s dummy model, perhaps because more happened during the 1980s than 1990s which impacted male longevity trends more than female longevity trends. The model with decade-specific terms for each decade other than the 1970s (which would be underidentified) performs best of all.

## Digression: reminding myself of the algebra and intuition of an asymptote

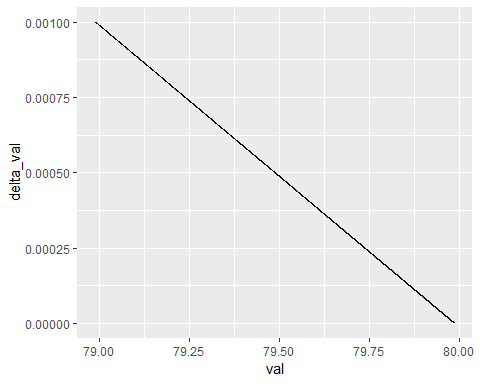
The improved fit with time suggests that maybe it would also be worth comparing a model with delta\_e0 with delta\_e0 squared, as the relationship might be best characterised as nonlinear rather than linear. To think through whether this should be the case when appraoching an asymptote, I should first produce a simple example of an asymptote

x <- 0:10000 / 1000  
K <- 80  
vals <- K - (1 / K + exp(-x))  
  
df <- data\_frame(  
 x = x,  
 val = vals,  
 delta\_val = val - lag(val)  
)   
  
ggplot(aes(x = x, y = val), data = df) +   
 geom\_line()



ggplot(aes(x = val, y = delta\_val), data = df) +  
 geom\_line()

## Warning: Removed 1 rows containing missing values (geom\_path).



This (of course) demonstrates through example something readily provable analytically: if life expectancy were approaching an asymptote then the coefficient of val ~ delta\_val will be a negative and nonlinear.

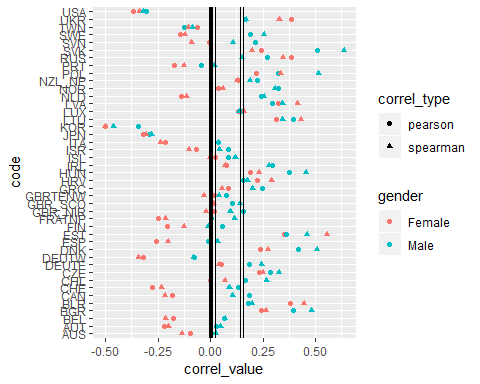
Given this, I’m not sure of the value of additionally adding a nonlinear term, as it doesn’t seem to be testing any kind of clear hypothesis. (Whereas regressing e0 on delta\_e0 clearly is)

**End of Digression!**

**Thinking again: it doesn’t quite make sense to add many time block factors together as this is controlling for something that should be inherent in the main model structure.**

However, I could simply show the correlation for each country

tmp %>%   
 filter(!code %in% c("DEUTNP", "FRACNP", "GBRCENW", "GBR\_NP", "NZL\_NM", "NZL\_MA")) %>%   
 group\_by(code, gender) %>%   
 nest() %>%   
 mutate(pearson = map\_dbl(data, ~cor(.x$e0, .x$delta\_e0)),  
 spearman = map\_dbl(data, ~cor(.x$e0, .x$delta\_e0, method = "spearman"))) %>%  
 gather(pearson:spearman, key = "correl\_type", value = "correl\_value") %>%   
 ggplot() +  
 geom\_point(aes(x = correl\_value, y = code, colour = gender, shape = correl\_type)) +   
 geom\_vline(xintercept = 0, size = 1.2) +   
 geom\_vline(aes(xintercept = mean(correl\_value[correl\_type == "pearson" & gender == "Female"]))) +   
 geom\_vline(aes(xintercept = mean(correl\_value[correl\_type == "pearson" & gender == "Male"]))) +  
 geom\_vline(aes(xintercept = mean(correl\_value[correl\_type == "spearman" & gender == "Female"]))) +  
 geom\_vline(aes(xintercept = mean(correl\_value[correl\_type == "spearman" & gender == "Male"])))



This clearly seems to show that there is no clear relationship internationally between life expectancy and change in life expectancy, and that the average correlation is around 0, or slightly positive. (around 0.1)