agData Vignette

Derek Wright

October 12, 2018

```
# devtools::install_github("derekmichaelwright/agData")
library(agData)
library(tidyverse)
```

Introduction

This is the vignette for the agData package.

```
?agData_FAO_Crops
?agData_FAO_Livestock
?agData_USDA_Crops
?agData_STATCAN_Crops
?agData_STATCAN_Livestock
?agData_STATCAN_Beehives
```

Load Data

A quick exploration of the data

```
# Load data
xx <- agData_FAO_Crops %>% as.tibble()
xx
```

```
## # A tibble: 2,157,696 x 6
##    Area        Item        Element        Unit    Year     Value
##    <fct>       <fct>       <fct>          <fct>  <dbl>     <dbl>
##  1 Afghanistan Apples      Area harvested ha      1961   2220
##  2 Afghanistan Apples      Yield          hg/ha   1961      6.80
##  3 Afghanistan Apples      Production     tonnes  1961  15100
##  4 Afghanistan Apricots    Area harvested ha      1961   4820
##  5 Afghanistan Apricots    Yield          hg/ha   1961      6.64
##  6 Afghanistan Apricots    Production     tonnes  1961  32000
##  7 Afghanistan Barley      Area harvested ha      1961 350000
##  8 Afghanistan Barley      Yield          hg/ha   1961      1.08
##  9 Afghanistan Barley      Production     tonnes  1961 378000
## 10 Afghanistan Berries nes Area harvested ha      1961   6800
## # ... with 2,157,686 more rows
```

```
# Spread data to wide format
xx %>%
  unite(Element, Element, Unit) %>%
  spread(Element, Value)
```

```
## # A tibble: 785,117 x 6
##    Area    Item       Year `Area harvested~ Production_tonn~ `Yield_hg/ha`
##    <fct>   <fct>     <dbl>            <dbl>            <dbl>         <dbl>
##  1 Afghan~ Almonds,~  1975                0                0         NA
##  2 Afghan~ Almonds,~  1976             5900             9800          1.66
```

```
##  3 Afghan~ Almonds,~  1977            6000             9000          1.5
##  4 Afghan~ Almonds,~  1978            6000            12000          2
##  5 Afghan~ Almonds,~  1979            6000            10500          1.75
##  6 Afghan~ Almonds,~  1980            5800             9900          1.71
##  7 Afghan~ Almonds,~  1981            5800             8000          1.38
##  8 Afghan~ Almonds,~  1982            5800            11000          1.90
##  9 Afghan~ Almonds,~  1983            5700             9700          1.70
## 10 Afghan~ Almonds,~  1984            5700            10500          1.84
## # ... with 785,107 more rows
```

```
# List measurements
xx %>% distinct(Element)
```

```
## # A tibble: 3 x 1
##   Element
##   <fct>
## 1 Area harvested
## 2 Yield
## 3 Production
```

```
# List areas
xx %>% distinct(Area)
```

```
## # A tibble: 258 x 1
##    Area
##    <fct>
##  1 Afghanistan
##  2 Albania
##  3 Algeria
##  4 American Samoa
##  5 Angola
##  6 Antigua and Barbuda
##  7 Argentina
##  8 Australia
##  9 Austria
## 10 Bahamas
## # ... with 248 more rows
```

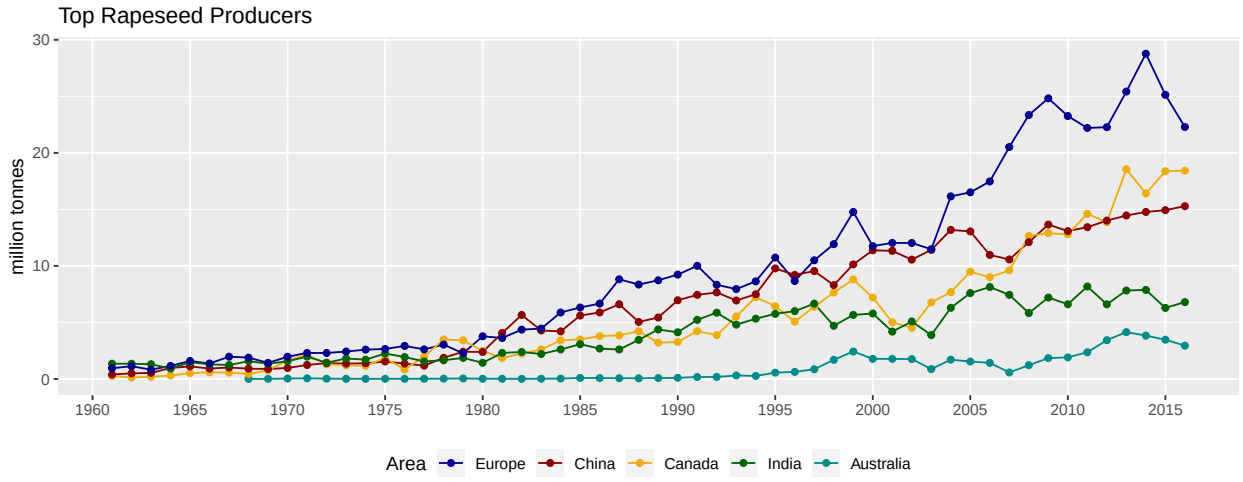
```
# List crops
xx %>% distinct(Item)
```

```
## # A tibble: 180 x 1
##    Item
##    <fct>
##  1 Apples
##  2 Apricots
##  3 Barley
##  4 Berries nes
##  5 Cotton lint
##  6 Cottonseed
##  7 Figs
##  8 Fruit, citrus nes
##  9 Fruit, fresh nes
## 10 Fruit, stone nes
## # ... with 170 more rows
```

Example 1: Rapeseed production

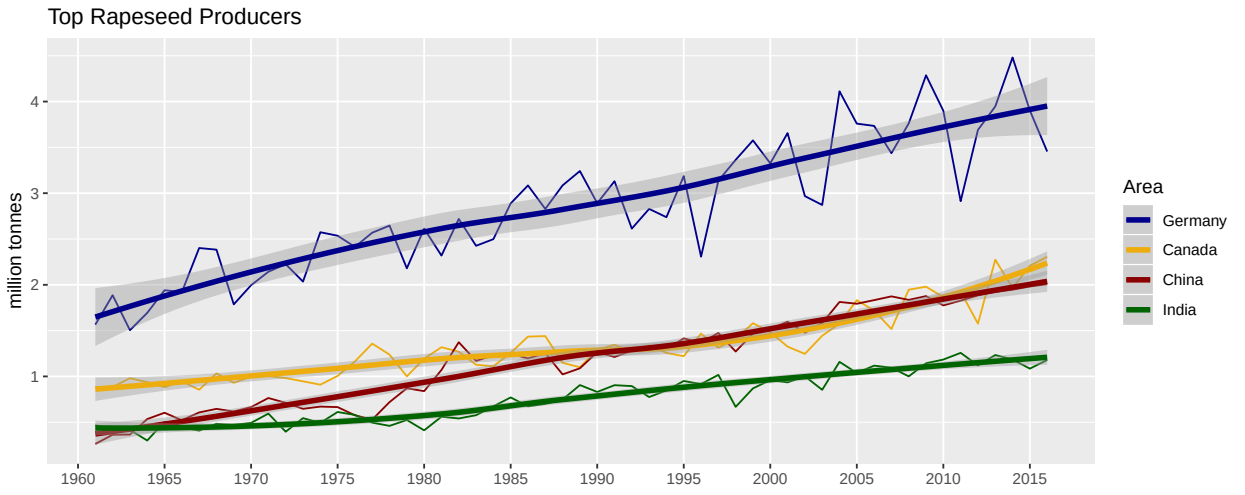
Improvments in oil quality, acheived through plant breeding has resulted in Rapeseed/Canola becoming one of the worlds major oil crops.

```
# Prep data
areas <- c("Europe", "China", "Canada", "India", "Australia")
cols  <- c("darkblue", "darkred", "darkgoldenrod2", "darkgreen", "darkcyan")
xx <- agData_FAO_Crops %>%
  filter(Item == "Rapeseed",
         Area %in% areas,
         Element == "Production") %>%
  mutate(Area = factor(Area, levels = areas))
# Plot
ggplot(xx, aes(x = Year, y = Value / 1000000, color = Area)) +
  geom_line() +
  geom_point() +
  scale_color_manual(values = cols) +
  scale_x_continuous(breaks       = seq(1960, 2015, by = 5),
                     minor_breaks = seq(1960, 2015, by = 5)) +
  theme(legend.position = "bottom") +
  labs(title = "Top Rapeseed Producers",
       y = "million tonnes", x = NULL)
```




```
# Prep data
areas <- c("Germany", "Canada", "China", "India")
cols  <- c("darkblue", "darkgoldenrod2", "darkred", "darkgreen")
xx <- agData_FAO_Crops %>%
  filter(Item == "Rapeseed",
         Area %in% areas,
         Element == "Yield") %>%
  mutate(Area = factor(Area, levels = areas))
# Plot
ggplot(xx, aes(x = Year, y = Value, color = Area)) +
  geom_line() +
  geom_smooth(method = "loess", size = 1.5) +
  scale_color_manual(values = cols) +
  scale_x_continuous(breaks       = seq(1960, 2015, by = 5),
                     minor_breaks = seq(1960, 2015, by = 5)) +
```

```
  labs(title = "Top Rapeseed Producers",
       y = "million tonnes", x = NULL)
```

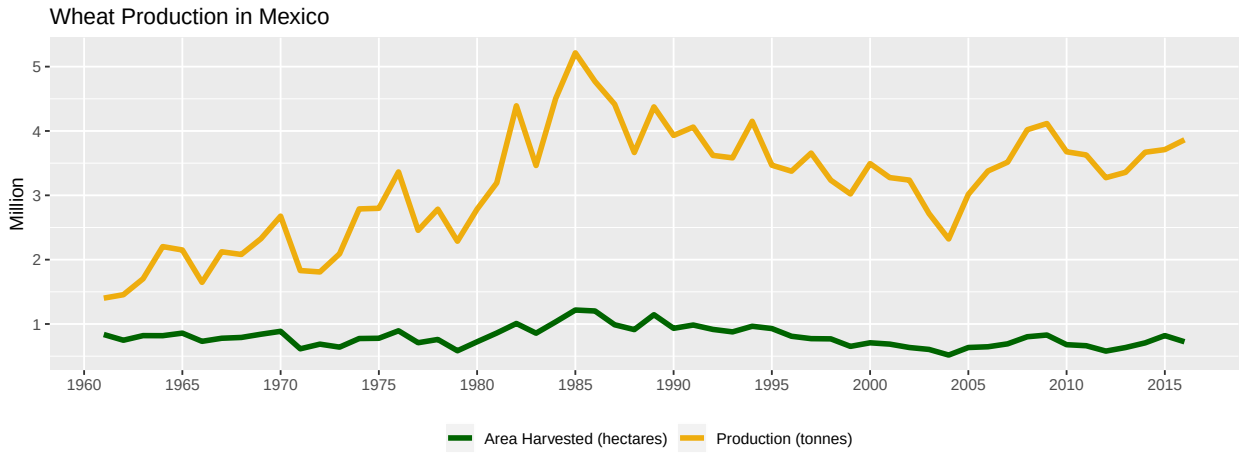


Example 2: Wheat production in India and Mexico

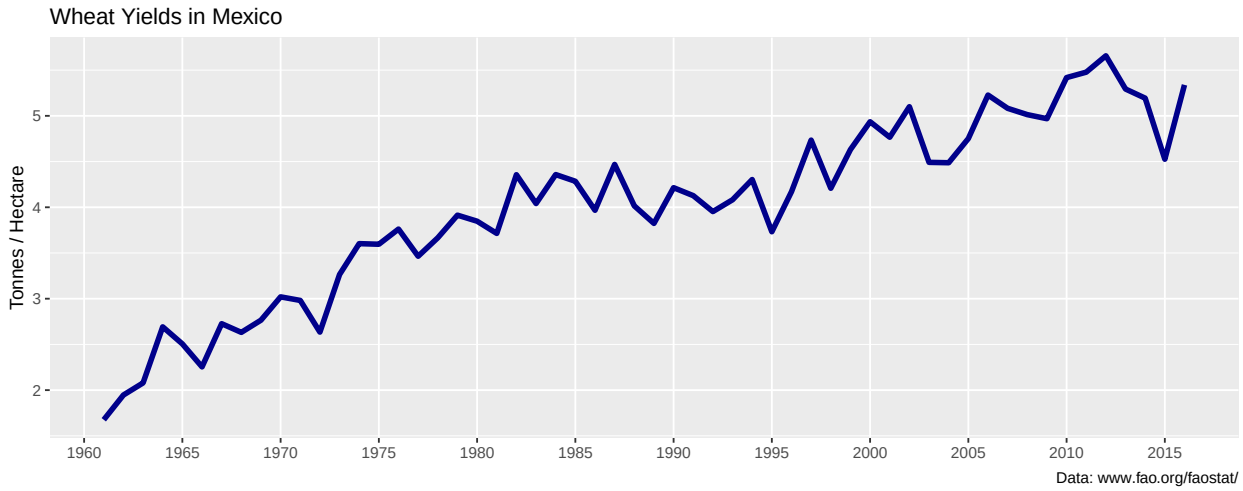
Spurred by pioneers such as Norman Bourlag, Wheat production in Mexico and India increased significantly since 1961. During that same time period, the area devoted to wheat production has remained relativly constant. This increase in wheat yields has helped these countries avoid some major food security problems.

```
# Prep data
xx <- agData_FAO_Crops %>%
  filter(Item == "Wheat",
         Area == "Mexico")
# Plot
ggplot(xx %>% filter(Element != "Yield"), aes(x = Year, y = Value / 1000000, color = Element)) +
  geom_line(size = 1.5) +
  theme(legend.position = "bottom") +
  scale_color_manual(name   = NULL,
                     labels = c("Area Harvested (hectares)", "Production (tonnes)"),
                     values = c("Dark Green", "darkgoldenrod2")) +
  scale_x_continuous(breaks       = seq(1960, 2015, by = 5),
                     minor_breaks = seq(1960, 2015, by = 5)) +

  labs(title   = "Wheat Production in Mexico",
       caption = "Data: www.fao.org/faostat/",
       y       = "Million",
       x       = NULL)
```



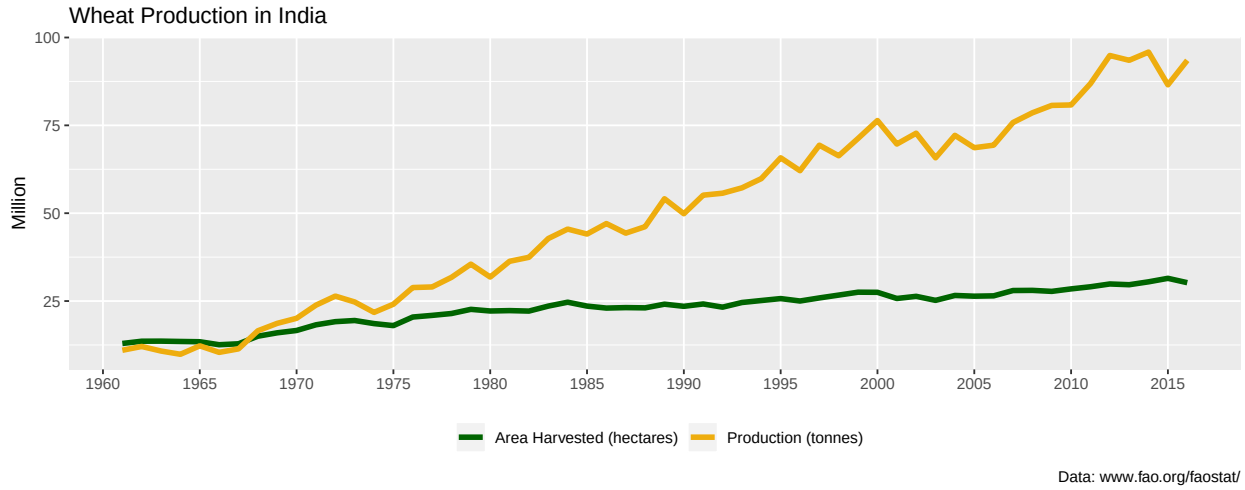

```
# Plot
ggplot(xx %>% filter(Element =="Yield"), aes(x = Year, y = Value) ) +
  geom_line(size = 1.5, color = "Dark Blue") +
  scale_x_continuous(breaks       = seq(1960, 2015, by = 5),
                     minor_breaks = seq(1960, 2015, by = 5)) +
  labs(title   = "Wheat Yields in Mexico",
       caption = "Data: www.fao.org/faostat/",
       y       = "Tonnes / Hectare",
       x       = NULL)
```



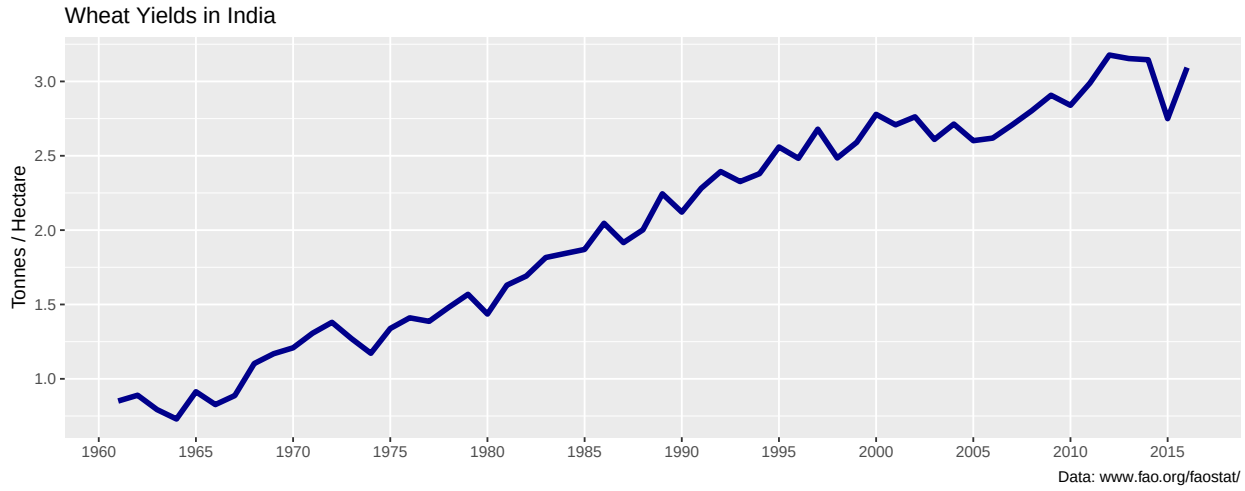

```
# Prep data
xx <- agData_FAO_Crops %>%
  filter(Item == "Wheat",
         Area == "India")
# Plot
ggplot(xx %>% filter(Element != "Yield"), aes(x = Year, y = Value / 1000000, color = Element)) +
  geom_line(size = 1.5) +
  theme(legend.position = "bottom") +
  scale_color_manual(name   = NULL,
                     labels = c("Area Harvested (hectares)", "Production (tonnes)"),
                     values = c("Dark Green", "darkgoldenrod2")) +
  scale_x_continuous(breaks       = seq(1960, 2015, by = 5),
```

```
                    minor_breaks = seq(1960, 2015, by = 5)) +

  labs(title   = "Wheat Production in India",
       caption = "Data: www.fao.org/faostat/",
       y       = "Million",
       x       = NULL)
```



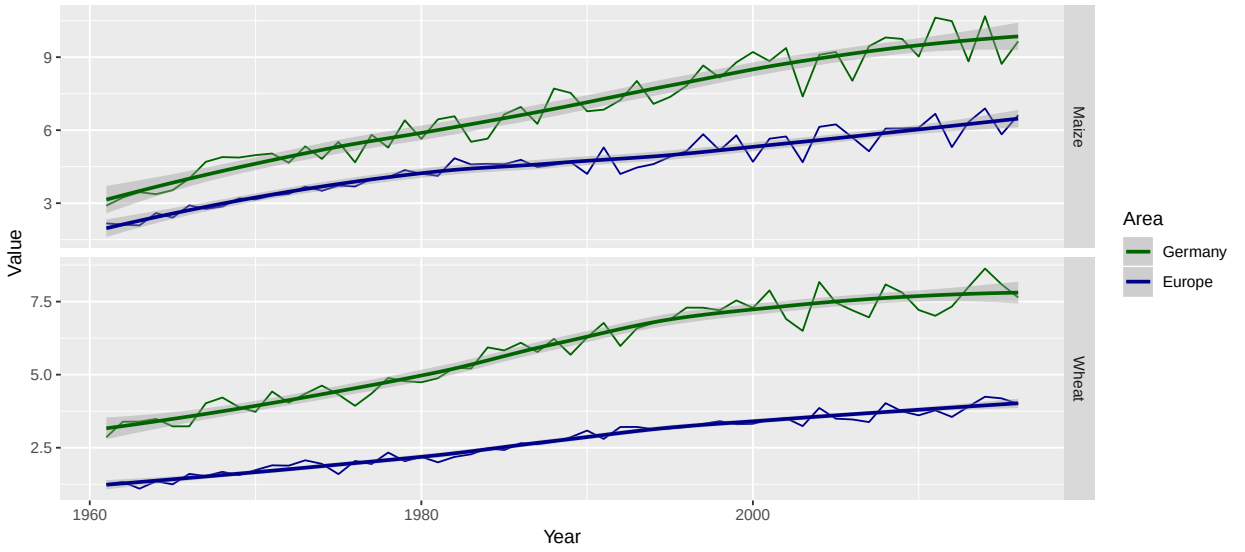

```
# Plot
ggplot(xx %>% filter(Element =="Yield"), aes(x = Year, y = Value) ) +
  geom_line(size = 1.5, color = "Dark Blue") +
  scale_x_continuous(breaks       = seq(1960, 2015, by = 5),
                     minor_breaks = seq(1960, 2015, by = 5)) +
  labs(title   = "Wheat Yields in India",
       caption = "Data: www.fao.org/faostat/",
       y       = "Tonnes / Hectare",
       x       = NULL)
```



Example 3: Wheat and Maize Yields in Germany vs Europe

Germany...

```r
# Prep data
xx <- agData_FAO_Crops %>%
  filter(Area %in% c("Germany", "Europe"),
         Item %in% c("Maize", "Wheat"),
         Element == "Yield") %>%
  mutate(Area = factor(Area, levels = c("Germany", "Europe")))
# Plot
ggplot(xx, aes(x = Year, y = Value, color = Area)) +
  geom_line() +
  geom_smooth(method = "loess") +
    facet_grid(Item~., scales = "free_y") +
  scale_color_manual(values = c("darkgreen", "darkblue"))
```

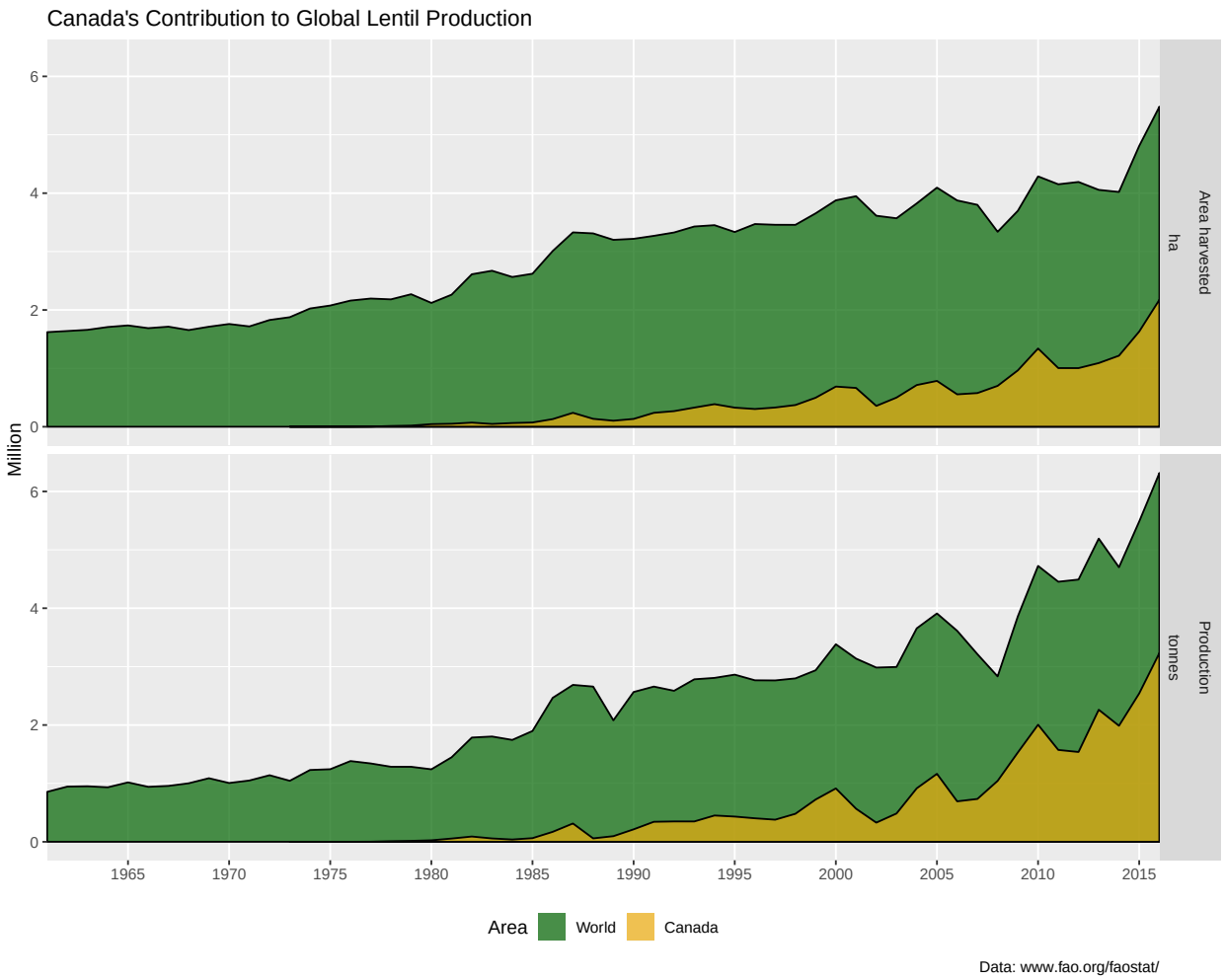


Example 4: FOA lentil data for Canada

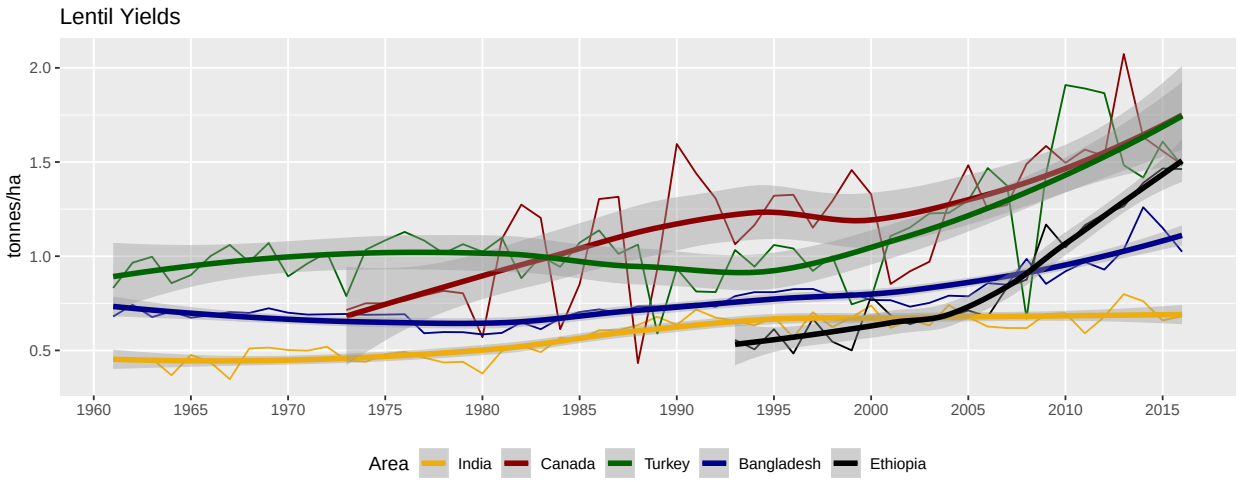
Since the introduction of lentil as a crop for the Canadian Prairies (1973), Saskatchewan has become the worlds largest producer of lentils. The first variety, "Laird", was registered in 1979.

```r
# Prep data
xx <- agData_FAO_Crops %>%
  filter(Item    == "Lentils",
         Element != "Yield",
         Area %in% c("Canada", "World") ) %>%
  mutate(Area = factor(Area, levels = c("World", "Canada")))
# Plot
ggplot(xx, aes(x = Year, y = Value / 1000000, fill = Area, color = I("Black"))) +
  geom_area(position = "identity", alpha = 0.7) +
  facet_grid(Element+Unit~.) +
  theme(legend.position = "bottom") +
  scale_fill_manual(values = alpha(c("Dark Green", "darkgoldenrod2"), 0.6)) +
  scale_x_continuous(breaks       = seq(1960, 2015, by = 5),
                     minor_breaks = seq(1960, 2015, by = 5),
                     expand = c(0,0))  +
```

```
  labs(title   = "Canada's Contribution to Global Lentil Production",
       caption = "Data: www.fao.org/faostat/",
       y = "Million", x = NULL)
```




```
# Prep data
areas <- c("India", "Canada", "Turkey", "Bangladesh", "Syria", "Ethiopia")
cols  <- c("darkgoldenrod2", "darkred", "darkgreen", "darkblue", "black", "darkcyan")
xx <- agData_FAO_Crops %>%
  filter(Item == "Lentils",
         Area %in% areas,
         Element == "Yield") %>%
  mutate(Area = factor(Area, levels = areas))
# Plot
ggplot(xx, aes(x = Year, y = Value, color = Area)) +
  geom_line() +
  geom_smooth(method = "loess", size = 1.5) +
  scale_x_continuous(breaks       = seq(1960, 2015, by = 5),
                     minor_breaks = seq(1960, 2015, by = 5)) +
  scale_color_manual(values = cols) +
  theme(legend.position = "bottom") +
  labs(title = "Lentil Yields",
       y = "tonnes/ha", x = NULL)
```

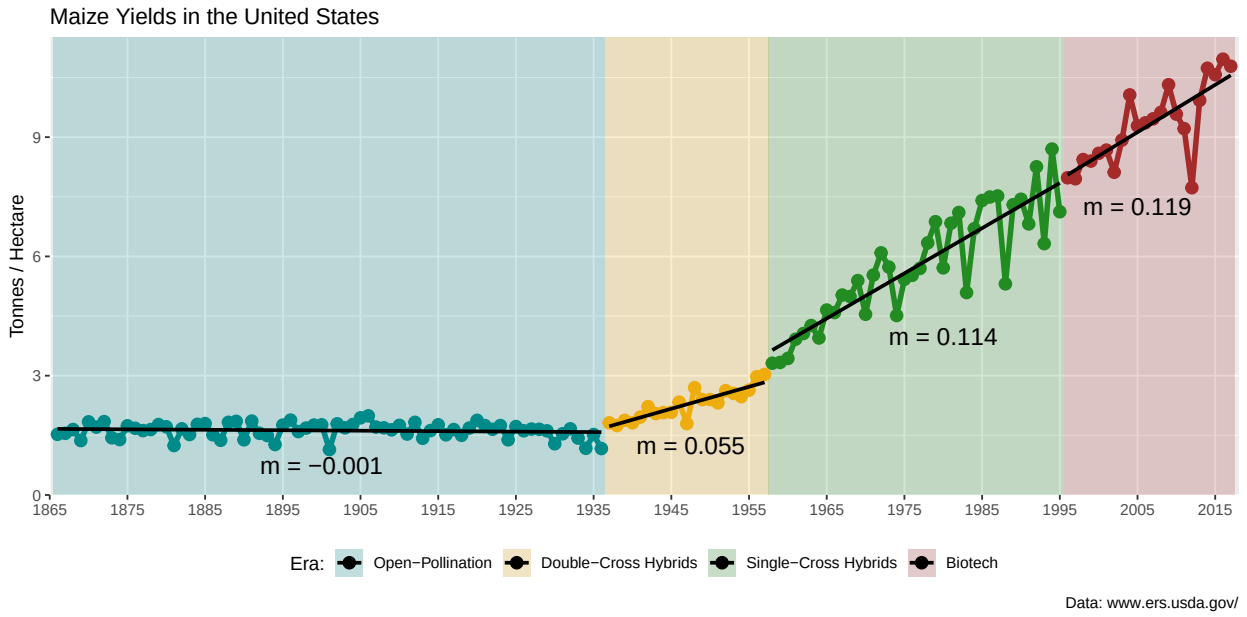


Example 5: USDA maize

The development of hybrid seed production in maize has led to major increases in crop yield in the United States.

```
# Prep data
xx <- agData_USDA_Crops %>%
  filter(Item == "Maize", Element == "Yield") %>%
  mutate(Era = ifelse(Year < 1937, "Open-Pollination",
                ifelse(Year < 1958, "Double-Cross Hybrids",
                 ifelse(Year < 1996, "Single-Cross Hybrids", "Biotech"))),
         Era = factor(Era, levels = c("Open-Pollination", "Double-Cross Hybrids",
                                      "Single-Cross Hybrids", "Biotech")))
# Prep rect data
x2 <- xx %>%
  group_by(Era) %>%
  summarise(min = min(Year), max = max(Year))
# Calculate slopes
c1 <- round(summary(lm(data = xx %>% filter(Era=="Open-Pollination"), Value~Year))$coefficients[2], 3)
c2 <- round(summary(lm(data = xx %>% filter(Era=="Double-Cross Hybrids"), Value~Year))$coefficients[2],
c3 <- round(summary(lm(data = xx %>% filter(Era=="Single-Cross Hybrids"), Value~Year))$coefficients[2],
c4 <- round(summary(lm(data = xx %>% filter(Era=="Biotech"), Value~Year))$coefficients[2], 3)
# Create color palette
cols <- c("darkcyan", "darkgoldenrod2", "Forest Green", "Brown")
# Plot
ggplot(xx, aes(fill = Era)) +
  geom_rect(data = x2, aes(xmin = min-0.5, xmax = max+0.5, ymin = -Inf, ymax = Inf), alpha = 0.2) +
  geom_line(size = 1.5, aes(x = Year, y = Value, color = Era)) +
  geom_point(size = 3, aes(x = Year, y = Value, color = Era)) +
  geom_smooth(method = "lm", se = F, colour = "Black", aes(x = Year, y = Value)) +
  scale_color_manual(name = "Era:", values = cols, guide = F) +
  scale_fill_manual(name = "Era:", values = cols) +
  scale_x_continuous(breaks       = seq(1865, 2015, by = 10),
                     minor_breaks = seq(1865, 2015, by = 10)) +
  coord_cartesian(xlim = c(1865, 2018), ylim = c(0, 11.5), expand = c(0, 0)) +
  annotate("text", x = 1900, y = 0.75, size = 5, label = paste("m =", c1)) +
```

```
  annotate("text", x = 1947.5, y = 1.25, size = 5, label = paste("m =", c2)) +
  annotate("text", x = 1980, y = 4, size = 5, label = paste("m =", c3)) +
  annotate("text", x = 2005, y = 7.25, size = 5, label = paste("m =", c4)) +
  theme(legend.position = "bottom") +
  labs(title   = "Maize Yields in the United States",
       caption = "Data: www.ers.usda.gov/",
       y       = "Tonnes / Hectare",
       x       = NULL)
```

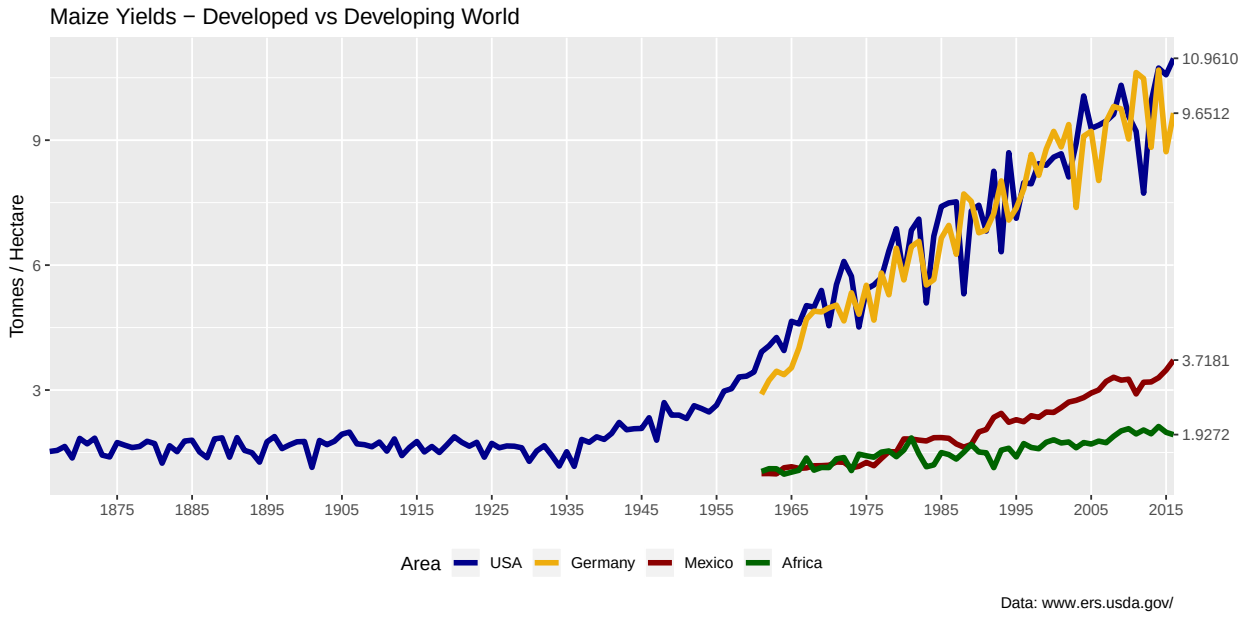


Example 6: Maize yields in the developed vs developing world

Maize yields in developing countries have lagged behind those in developed countries. This is due to a conbination of factors, including lack of access to crop inputs, machinery, and improved crop varieties.

```
# Prep data
x1 <- agData_USDA_Crops %>%
  filter(Item    == "Maize",
         Element == "Yield",
         Year != 2017)
x2 <- agData_FAO_Crops %>%
  filter(Item    == "Maize",
         Element == "Yield",
         Area %in% c("Germany", "Mexico", "Africa"))
xx <- bind_rows(x1, x2) %>%
  mutate(Area = factor(Area, levels = c("USA", "Germany", "Mexico", "Africa")))
xE <- xx %>% top_n(1, Year) %>% pull(Value)
# Plot
ggplot(xx, aes(x = Year, y = Value, color = Area)) +
  geom_line(size = 1.5) +
  theme(legend.position = "bottom") +
  scale_color_manual(values = c("Dark Blue", "darkgoldenrod2", "Dark Red", "Dark Green")) +
```

```
  scale_x_continuous(breaks       = seq(1865, 2015, by = 10),
                     minor_breaks = seq(1865, 2015, by = 10),
                     expand = c(0, 0)) +
  scale_y_continuous(sec.axis = sec_axis(~ ., breaks = xE)) +
  labs(title    = "Maize Yields - Developed vs Developing World",
       caption = "Data: www.ers.usda.gov/",
       y       = "Tonnes / Hectare",
       x       = NULL)
```

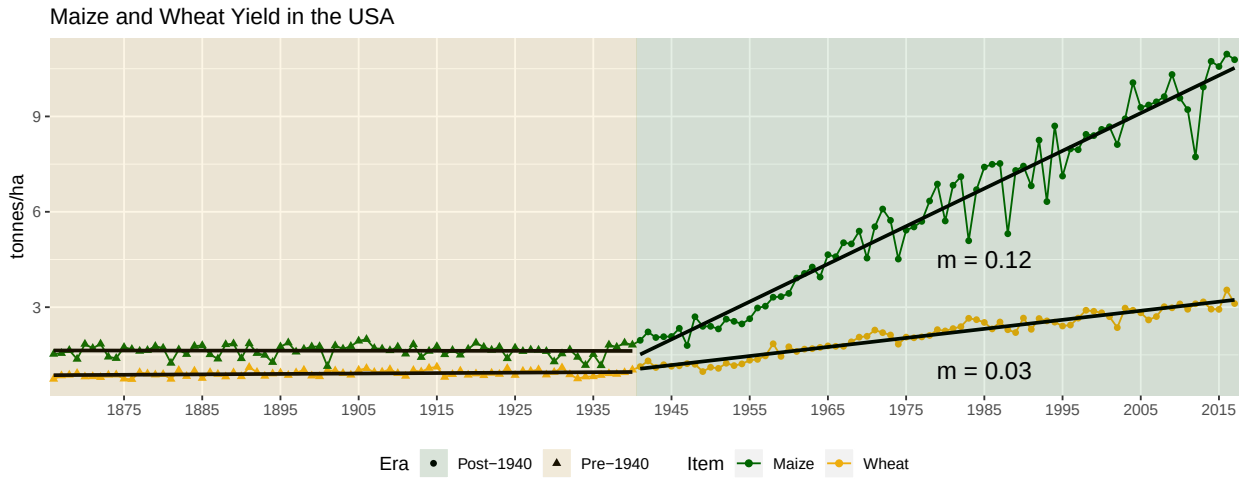


Example 7: USDA Maize vs Wheat yields

Maize yeilds have increased on a much faster pace than wheat, in part due to the adoption of hybrid seed in Maize.

```
# Prep data
xx <- agData_USDA_Crops %>%
  filter(Item %in% c("Maize", "Wheat"),
         Element == "Yield") %>%
  mutate(Era = ifelse(Year <= 1940, "Pre-1940", "Post-1940"))
#
x2 <- xx %>%
  group_by(Era) %>%
  summarise(min = min(Year), max = max(Year))
#
c1 <- round(summary(lm(data = xx %>% filter(Item=="Maize",Era=="Post-1940"), Value~Year))$coefficients[
c2 <- round(summary(lm(data = xx %>% filter(Item=="Wheat",Era=="Post-1940"), Value~Year))$coefficients[
# Plot
ggplot(xx) +
  geom_line(aes(x = Year, y = Value, color = Item)) +
  geom_point(aes(x = Year, y = Value, color = Item, shape = Era)) +
  geom_smooth(data = xx %>% filter(Item == "Wheat"),
              method = "lm", se = F, colour = "Black", aes(x = Year, y = Value, group = Era)) +
```

```
  geom_smooth(data = xx %>% filter(Item == "Maize"),
              method = "lm", se = F, colour = "Black", aes(x = Year, y = Value, group = Era)) +
  geom_rect(data = x2, aes(xmin = min-0.5, xmax = max+0.5, ymin = -Inf, ymax = Inf, fill = Era), alpha =
  annotate("text", x = 1985, y = 4.5, size = 5, label = paste("m =", c1)) +
  annotate("text", x = 1985, y = 1, size = 5, label = paste("m =", c2)) +
  scale_x_continuous(breaks       = seq(1865, 2015, by = 10),
                     minor_breaks = seq(1865, 2015, by = 10),
                     expand = c(0,0)) +
  scale_color_manual(values = c("darkgreen", "darkgoldenrod2")) +
  scale_fill_manual(values = c("darkgreen", "darkgoldenrod2")) +
  theme(legend.position = "bottom") +
  labs(title = "Maize and Wheat Yield in the USA",
       y = "tonnes/ha", x = NULL)
```

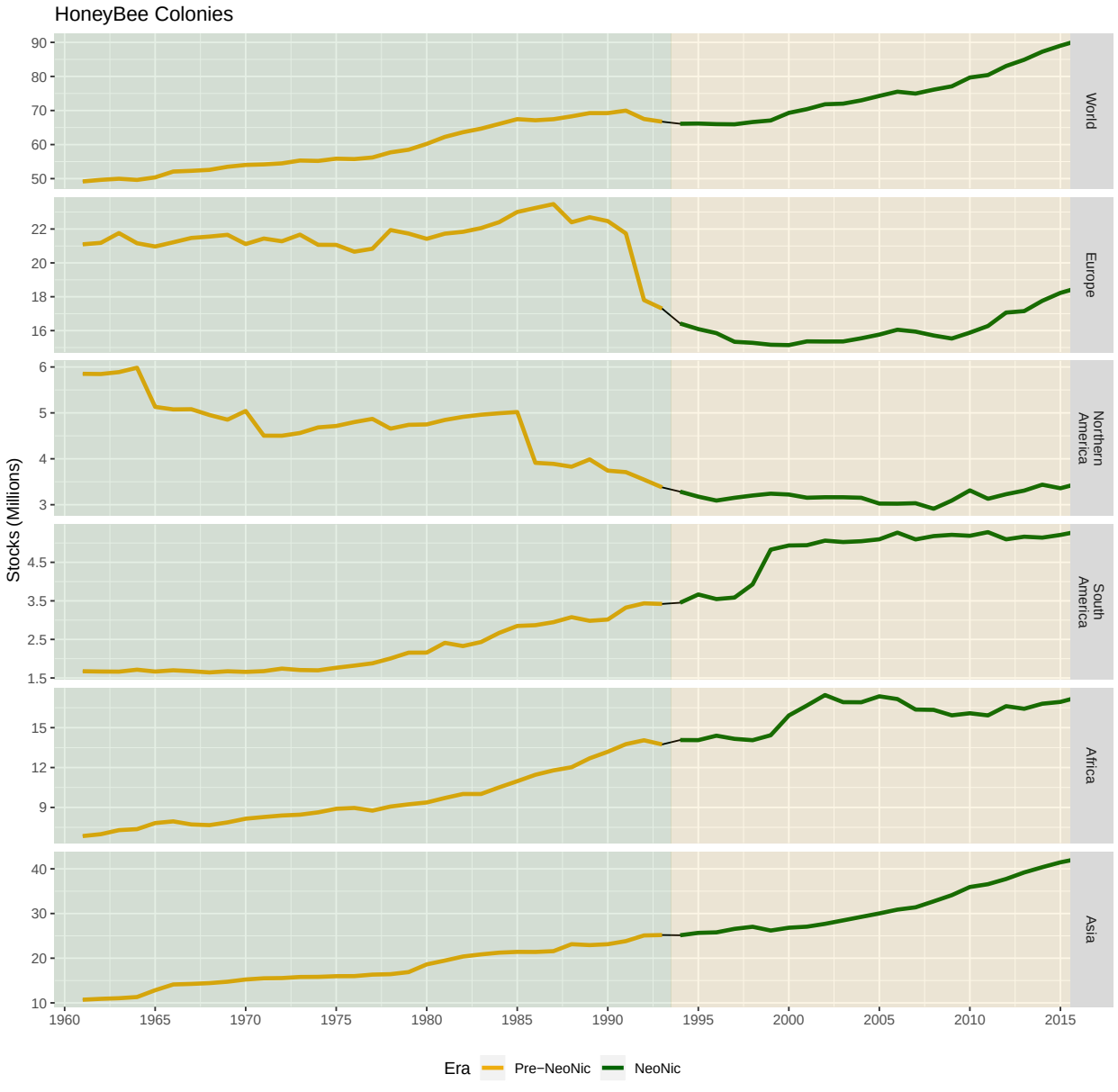


Example 8: FAO and STATCAN honeybee data

Neonicotinoids are often blamed for honeybee declines, but the data suggests a more complicated story.

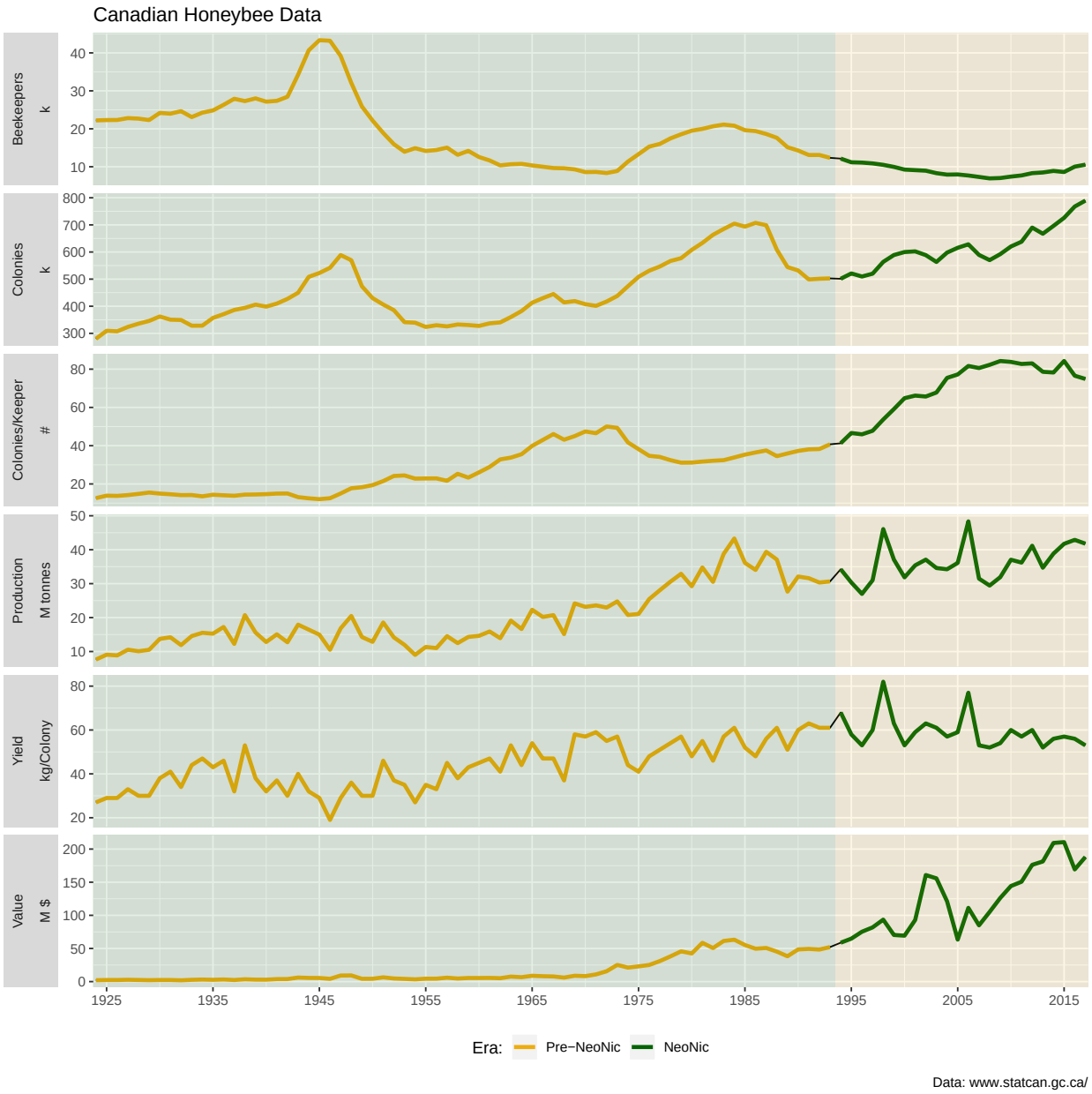
```
# Prep data
areas <- c("World", "Europe", "Northern America", "South America", "Africa", "Asia")
xx <- agData_FAO_Livestock %>%
  filter(Item == "Beehives") %>%
  mutate(Era = ifelse(Year >= 1994, "NeoNic", "Pre-NeoNic"),
         Era = factor(Era, levels = c("Pre-NeoNic", "NeoNic") ) ) %>%
  filter(Area %in% areas) %>%
  mutate(Area = factor(Area, levels = areas))
# Plot
ggplot(data = xx, aes(x = Year, y = Value / 1000000)) +
  geom_line() + geom_line(aes(color = Era), size = 1.25) +
  facet_grid(Area ~ ., scales = "free_y", labeller = label_wrap_gen(width = 10)) +
  theme(legend.position = "bottom") +
  scale_color_manual(values = c("darkgoldenrod2", "Dark Green")) +
  scale_x_continuous(breaks = seq(1960, 2015, by = 5)) +
  coord_cartesian(xlim = c(1962, 2013)) +
  annotate("rect", xmin = 1940, xmax = 1993.5, ymin = -Inf, ymax = Inf, fill = "Dark Green",      alpha =
```

```
  annotate("rect", xmin = 1993.5, xmax = 2035, ymin = -Inf, ymax = Inf, fill = "darkgoldenrod2", alpha =
  labs(title = "HoneyBee Colonies",
       caption = "Data: www.fao.org/faostat/",
       y = "Stocks (Millions)",
       x = NULL)
```



```
# Prep data
levs <- c("Beekeepers", "Colonies", "Colonies/Keeper",  "Production", "Yield", "Value")
xx <- agData_STATCAN_Beehives %>%
  filter(Area == "Canada") %>%
  select(-Unit) %>%
  spread(Element, Value) %>%
```

```
  mutate(Colonies = Colonies / 1000,
         Production = Production / 1000000,
         Beekeepers = Beekeepers / 1000) %>%
  gather("Element", "Value", Colonies, Production, Beekeepers, ColoniesPerBeekeeper, Yield, Value) %>%
  mutate(Element = plyr::mapvalues(Element, "ColoniesPerBeekeeper", "Colonies/Keeper"),
         Element = factor(Element, levels = levs),
         Unit = plyr::mapvalues(Element, levs, c("k","k","#","M tonnes","kg/Colony","M $")),
         Era = ifelse(Year >= 1994, "NeoNic", "Pre-NeoNic"),
         Era = factor(Era, levels = c("Pre-NeoNic", "NeoNic")))
#
ggplot(xx, aes(x = Year, y = Value)) +
  geom_line() +
  geom_line(aes(color = Era), size = 1.25) +
  facet_grid(Element + Unit ~ ., scales = "free", switch = "y") +
  scale_colour_manual(name = "Era:", values = c("darkgoldenrod2", "Dark Green")) +
  scale_x_continuous(breaks = seq(1925, 2015, by = 10)) +
  coord_cartesian(xlim = c(min(xx$Year)+4, max(xx$Year)-4)) +
  annotate("rect", xmin = 1905, xmax = 1993.5, ymin = -Inf, ymax = Inf, fill = "Dark Green",     alpha =
  annotate("rect", xmin = 1993.5, xmax = 2035, ymin = -Inf, ymax = Inf, fill = "darkgoldenrod2", alpha =
  theme(strip.placement = "outside", legend.position = "bottom", axis.title.y = element_text(hjust = 0))
  labs(title = "Canadian Honeybee Data",
       caption = "Data: www.statcan.gc.ca/",
       y = NULL, x = NULL)
```



Example 9: gganimate example

```
library(gganimate)
# Prep data
xx <- agData_STATCAN_Beehives %>%
  filter(Area    == "Canada",
         Element == "Colonies")
# Create GIF
mp <- ggplot(xx, aes(Year, Value / 1000, group = Area)) +
  geom_line() +
  labs(title = "Honeybee Colonies in Canada - {frame_along}",
       caption = "Data: www.statcan.gc.ca/",
```

```
       y = "Thousand Colonies") +
  transition_reveal(Area, Year)
mp <- animate(mp, fps = 5)
anim_save("anim.gif", mp)
```

HexSticker Creation

```
library(hexSticker)
# Prep data
xx <- agData_FAO_Crops %>%
  filter(Area    == "Canada",
         Item    == "Lentils",
         Element == "Production")
# Create sticker
mp <- ggplot(xx, aes(x = Year, y = Value/1000000)) +
  geom_line() + theme(axis.text = element_blank(), axis.ticks = element_blank()) +
  labs(x = NULL, y = NULL) +
  theme_transparent()
sticker(mp,
        package="agData", p_color = "Black", p_size = 30,
        s_x = 1, s_y = 0.8, s_width = 1.75, s_height = 0.8,
        h_fill = "#614105", h_color = "darkolivegreen", h_size = 3)
```