Google Data Analytics Professional Certificate - Capstone Project:

European Engine Trend Analysis 2001 - 2019

Jefferson Galvez | https://www.linkedin.com/in/jefferson-galvez/

2022-11-06

Introduction and Data Source

In 2018, the total number of registered passenger cars in Europe was 292 million [1], and at the start of that year, the European Union (EU) alone had a population of almost 513 million [2]. That's approximately 1 passenger car per every 1.755 European within the EU alone, and there appears to be no signs of decline in either population or economic growth which would support the idea of decreasing vehicular registration, despite Europe's extensive public transportation system. This phenomenon presents opportunity and growth for many automotive companies, to further develop their cars and expand their revenue. However, with growing concerns regarding environmental sustainability and development of electric car technology. The automotive industry is faced with a unique challenge to fulfil their moral obligation to protect our ecosystems, yet deliver transportation means to an ever growing world.

This case study will investigate the recent trends of vehicular engine types (diesel, petrol, battery electric, plug-in hybrid and full mild hybrids) in Europe from 2001 to 2019. Looking into the most optimal engine type to focus production and sales, to maximize an automotive company's growth. Assessing historical trends and creating predictions.

Data sources

1. Dataset Reference: Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "Energy". https://ourworldindata.org/transport

2. GDP data source: International Monetary Fund (IMF) International Monetary Fund, WEO Database October 2009

Phase 1: Ask

Business task

- Identify any trends in new passenger vehicular registrations engine types in Europe between 2001 to 2019
- Make predictions on future trending engine types based on any popular/'winning' engine types

Stakeholder and audience

The stakeholders would include the executive members of the following departments: sales, marketing, business development, production and product improvement. All such members would utilize the insights from this analysis to make data-driven business decisions, focusing efforts on promising products, understanding market demands and tailoring business processes to adapt for the future market.

Phase 2: Prepare

Data description

The dataset containing information on new vehicular engine type registrations, in Europe from 2001 to 2019 was downloaded as a .csv and uploaded onto Google Drive and a local MySQL server. The file was named ‘new-vehicles-type-area’.

The dataset contains the count of each new passenger vehicular registrations engine type, in specific European countries. Only one dataset from the IMF was used in this project to refer to the annual GDP (nominal), for each European country between 2001 to 2019, mentioned above under ‘Introduction and Data Source’.

Data credibility

Does the data ROCCC?

Reliable: Regarding vehicular engine type data, OurWorldInData sources their data from the International Council on Clean Transport (ICCT) and European Environment Agency. The ICCT is independent nonprofit organization (NGO), with a purpose to provide unbiased research, technical and scientific analysis to environmental regulators. The IMF comprises of 190 sovereign nations which supports economic growth and financial cooperation, amongst its members. They are governed and accountable by their own member nations. The founding purpose of the ICCT grants reliable credibility to the data obtained from OurWorldInData. The large number of member states within the IMF, would lead to fair accountability for all, therefore also granting reliable credibility for the GDP data obtained.

Original: Nominal GDP data from the IMF is original, however OurWorldInData does not state whether or not the data has been processed in any manner.

Comprehensive: All required data regarding engine types and nominal GDP is included from both data sources, with the exception of missing engine data from Iceland from 2001 - 2015, 2018 and 2019.

Current: As of the beginning of this project (Q1 2022), the latest available data on engine from OurWorldInData is not the most current only going up to 2019. The article was first published in September 2021 with no mention of missing 2020 and 2021 data. One can assume this may be attributed to the COVID-19 pandemic. However, given the longitudinal analysis nature of this project, the data date range is acceptable.

Cited: Both sources are cited.

Phase 3: Process

Data Cleaning, Processing and Storage via MySQL

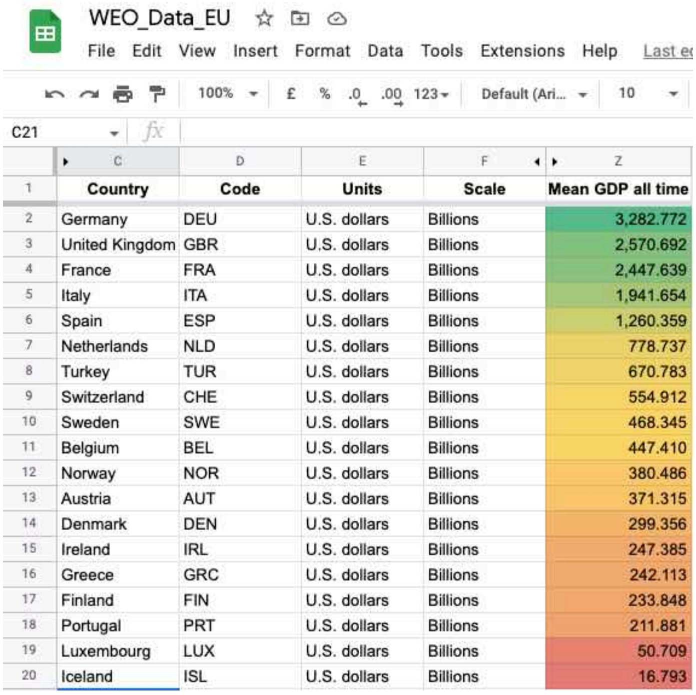
Three data sets stored under ‘engine_types_project’ database in MySQL. Total of 1,236 rows of data was transferred. Due to the sheer size of data available, the project will shift in analyzing only the top 3 European countries by nominal GDP

MySQL data processing

```
- Viewed every unique Entity (country) in the new-vehicles-
type-area table to observe which countries are involved in
this project

- 20 Countries/rows returned, same was true for the other
three tables

SELECT DISTINCT entity
FROM new-vehicles-type-area


- Renamed the column Entity to Country for all tables, for ease
of understanding

ALTER TABLE new-vehicles-type-area
RENAME COLUMN Entity TO Country;

- Observed that Europe was listed as a Country within the
Country Column, therefore, rows containing Europe under the
Country field was deleted from all tables

SET SQL_SAFE_UPDATES = 0;
DELETE FROM new-vehicles-type-area
WHERE Country = 'Europe';
SET SQL_SAFE_UPDATES = 1;


- Displaying and counting all unique countries within working
data set and their code

SELECT COUNT(DISTINCT Country
FROM new-vehicles-type-area AS
Country_Number; SELECT Country, Code
FROM new-vehicles-type-area
GROUP BY  Country, Code;
```

Using Google Sheets to rank European countries by GDP (nominal)

- GDP of each country was averaged across the years and sheets was sorted by the average column in descending order
- Average column formatted with colour scale to further emphasize GDP ranking



	Country	Code	Units	Scale	Mean GDP all time
2	Germany	DEU	U.S. dollars	Billions	3,282.772
3	United Kingdom	GBR	U.S. dollars	Billions	2,570.692
4	France	FRA	U.S. dollars	Billions	2,447.639
5	Italy	ITA	U.S. dollars	Billions	1,941.654
6	Spain	ESP	U.S. dollars	Billions	1,260.359
7	Netherlands	NLD	U.S. dollars	Billions	778.737
8	Turkey	TUR	U.S. dollars	Billions	670.783
9	Switzerland	CHE	U.S. dollars	Billions	554.912
10	Sweden	SWE	U.S. dollars	Billions	468.345
11	Belgium	BEL	U.S. dollars	Billions	447.410
12	Norway	NOR	U.S. dollars	Billions	380.486
13	Austria	AUT	U.S. dollars	Billions	371.315
14	Denmark	DEN	U.S. dollars	Billions	299.356
15	Ireland	IRL	U.S. dollars	Billions	247.385
16	Greece	GRC	U.S. dollars	Billions	242.113
17	Finland	FIN	U.S. dollars	Billions	233.848
18	Portugal	PRT	U.S. dollars	Billions	211.881
19	Luxembourg	LUX	U.S. dollars	Billions	50.709
20	Iceland	ISL	U.S. dollars	Billions	16.793

Top 3 European countries by GDP (nominal):

1. Germany
2. United Kingdom
3. France

- The European GDP data was imported into MySQL workbench as .csv file
- More cleaning and processing of the GDP data, from the IMF, was required due to data type differences upon importing
- Datatype conversion required for proper future analysis and merging with engine data

```
- Data type was imported as TEXT (due to .csv format),
therefore data type needed to be converted into DECIMAL()

- MySQL would truncate data which contained commas, therefore
commas were removed from each column containing numeric data
prior to data type conversion

update weo_data_eu set `Mean GDP all time` = replace(`Mean GDP
all time`,',','');

update weo_data_eu set `2001` = replace(`2001`,',','');

update weo_data_eu set `2002` = replace(`2002`,',','');

update weo_data_eu set `2003` = replace(`2003`,',','');

update weo_data_eu set `2004` = replace(`2004`,',','');

update weo_data_eu set `2005` = replace(`2005`,',','');

update weo_data_eu set `2006` = replace(`2006`,',','');
```

```
update weo_data_eu set `2007` = replace(`2007`,',','');

update weo_data_eu set `2008` = replace(`2008`,',','');

update weo_data_eu set `2009` = replace(`2009`,',','');

update weo_data_eu set `2010` = replace(`2010`,',','');

update weo_data_eu set `2011` = replace(`2011`,',','');

update weo_data_eu set `2012` = replace(`2012`,',','');

update weo_data_eu set `2013` = replace(`2013`,',','');

update weo_data_eu set `2014` = replace(`2014`,',','');

update weo_data_eu set `2015` = replace(`2015`,',','');

update weo_data_eu set `2016` = replace(`2016`,',','');

update weo_data_eu set `2017` = replace(`2017`,',','');

update weo_data_eu set `2018` = replace(`2018`,',','');

update weo_data_eu set `2019` = replace(`2019`,',','');

– Data type within each year column was converted into
DECIMAL(6,2)

ALTER TABLE `engine_type_project`.`weo_data_eu`
CHANGE COLUMN `Mean GDP all time` `Mean GDP all time`
DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu`
CHANGE COLUMN `2001` `2001` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2002` `2002` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2003` `2003` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2004` `2004` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2005` `2005` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2006` `2006` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2007` `2007` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2008` `2008` DECIMAL(6,2) NULL DEFAULT NULL;
```

```
ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2009` `2009` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2010` `2010` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2011` `2011` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2012` `2012` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2013` `2013` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2014` `2014` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2015` `2015` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2016` `2016` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2017` `2017` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2018` `2018` DECIMAL(6,2) NULL DEFAULT NULL;

ALTER TABLE `engine_type_project`.`weo_data_eu` CHANGE COLUMN
`2019` `2019` DECIMAL(6,2) NULL DEFAULT NULL;

– Creating tables for top 3 by using the ORDER BY clause, it
was simple to gather the relevant top and bottom 3 European
countries from the same table, by reversing the order – Only
the ‘Code’ and individual year columns were kept for both
tables for ease of understanding and future merging with
engine tables

CREATE TABLE Top_3_EU_GDP ( SELECT * FROM weo_data_eu ORDER BY
Mean GDP all time DESC LIMIT 3 );

– Dropping irrelevant columns for ease of table use and data
interpretation
ALTER TABLE Top_3_EU_GDP
DROP COLUMN WEO Country Code,
DROP COLUMN WEO Subject Code,
DROP COLUMN Mean GDP all time,
DROP COLUMN Country,
DROP COLUMN Units,
DROP COLUMN Scale;
```

Connecting MySQL to R

- MySQL database was connected to RStudio for seamless data processing and future analysis within RStudio
- Connection will guarantee I am working on the most update data, should the data be updated from MySQL

```
# loading packages
install.packages('tidyverse', repos = "http://cran.us.r-
project.org")

##
## The downloaded binary packages are in
##   /var/folders/s4/1njj99vx55bgp7h4b35gmwk40000gn/T//Rtmpdw6cEl/
downloaded_packages

library(tidyverse)

## ── Attaching packages ───────────────────────────────────────
tidyverse 1.3.2 ──
## ✔ ggplot2 3.4.0      ✔ purrr   0.3.4
## ✔ tibble  3.1.8      ✔ dplyr   1.0.10
## ✔ tidyr   1.2.0      ✔ stringr 1.4.0
## ✔ readr   2.1.2      ✔ forcats 0.5.1
## ── Conflicts ──────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()

install.packages('cowplot', repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##   /var/folders/s4/1njj99vx55bgp7h4b35gmwk40000gn/T//Rtmpdw6cEl/
downloaded_packages

library(cowplot)
install.packages('scales', repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##   /var/folders/s4/1njj99vx55bgp7h4b35gmwk40000gn/T//Rtmpdw6cEl/
downloaded_packages

library(scales)

##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
```

```
## The following object is masked from 'package:readr':
##
##     col_factor

install.packages('datasets', repos = "http://cran.us.r-project.org")

## Warning: package 'datasets' is a base package, and should not be
updated

library(datasets)
install.packages('ggplot2', repos = "http://cran.us.r-project.org")

##
##   There is a binary version available but the source version is
later:
##         binary source needs_compilation
## ggplot2  3.3.6  3.4.0             FALSE

## installing the source package 'ggplot2'

library(ggplot2)
install.packages('dplyr', repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##  /var/folders/s4/1njj99vx55bgp7h4b35gmwk40000gn/T//Rtmpdw6cEl/
downloaded_packages

library(dplyr)
install.packages('DBI', repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##  /var/folders/s4/1njj99vx55bgp7h4b35gmwk40000gn/T//Rtmpdw6cEl/
downloaded_packages

library(DBI)
install.packages('RMariaDB', repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##  /var/folders/s4/1njj99vx55bgp7h4b35gmwk40000gn/T//Rtmpdw6cEl/
downloaded_packages

library(RMariaDB)

# specify parameters
user <- "root"
password <- "[redacted]"
database <-  "engine_type_project"
host <- "Jeffersons-MacBook-Air.local" # local host
port <- 3306

# connect to a database
```

```
connection <- dbConnect(
  drv = MariaDB(),
  dbname = database,
  user = user,
  password = password,
  port = port
)

# fetch results
tbl(connection, "new-vehicles-type-area") %>%
  collect() -> type_by_area
tbl(connection, "Top_3_type_area") %>%
  collect() -> top_3_area
tbl(connection, "Top_3_EU_GDP") %>%
  collect() -> Top_3
```

Converting wide data to long data

In the context of this analysis, it makes more logical sense for the GDP table to be re-shaped into long data

```
# wide data to be converted into long
Top_3

## # A tibble: 3 × 20
##   Code  `2001` `2002` `2003` `2004` `2005` `2006` `2007` `2008`
`2009` `2010`
##   <chr>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
<dbl>  <dbl>
## 1 DEU    1946.  2077.  2501.  2813.  2848.  2995.  3426.  3745.
3408.  3402.
## 2 GBR    1626.  1776.  2046.  2405.  2528.  2701.  3085.  2935.
2403.  2455.
## 3 FRA    1378.  1500.  1844.  2119.  2198.  2321.  2661.  2930.
2698.  2647.
## # … with 9 more variables: `2011` <dbl>, `2012` <dbl>, `2013`
<dbl>,
## #   `2014` <dbl>, `2015` <dbl>, `2016` <dbl>, `2017` <dbl>,
`2018` <dbl>,
## #   `2019` <dbl>

# converting from wide to long data
Top_3_long <- Top_3 %>%
  gather(Year, GDP, -c(Code))
```

Merging tables

- Merged GDP table with ‘top_3_area’ tables

- Creates tables which displays the engine type count for each country, for each year as well as GDP (nominal) for that specific year

```
# merging top 3 long data with top 3 area data
top_3_gdp_area <- merge(top_3_area, Top_3_long, by = c("Year",
"Code"))
options(max.print=1000000)  # setting high maximum print limit to
prevent table truncation
```

Creating country-specific data frames

Tree maps of Germany, France and Great Britain will be required later for analysis. For this, country-specific data frames are required in a long format with only the necessary columns.

```
# Subsetting country data
# Germany
DEU_2001_df <- top_3_gdp_area[top_3_gdp_area$Code == 'DEU',] %>%
slice(1)
DEU_2019_df <- top_3_gdp_area[top_3_gdp_area$Code == 'DEU',] %>%
slice(19)

# France
FRA_2001_df <- top_3_gdp_area[top_3_gdp_area$Code == 'FRA',] %>%
slice(1)
FRA_2019_df <- top_3_gdp_area[top_3_gdp_area$Code == 'FRA',] %>%
slice(19)

#Great Britain
GBR_2001_df <- top_3_gdp_area[top_3_gdp_area$Code == 'GBR',] %>%
slice(1)
GBR_2019_df <- top_3_gdp_area[top_3_gdp_area$Code == 'GBR',] %>%
slice(19)


# Recreating table as text to convert data from wide to long,
dropping columns X and GDP
DEU_2001_df <- read.table(text = "        Year petrol_number
diesel_gas_number full_mild_hybrid_number plugin_hybrid_number
battery_electric_number
                          1    2001 2185247 1155300 606.0005 0 56",
header = TRUE, sep = "")
DEU_2019_df <- read.table(text = "        Year petrol_number
diesel_gas_number full_mild_hybrid_number plugin_hybrid_number
battery_electric_number
                          1    2019 2120699 1222769 155461.8 44640
63461", header = TRUE, sep = "")

FRA_2001_df <- read.table(text = "        Year petrol_number
diesel_gas_number full_mild_hybrid_number plugin_hybrid_number
battery_electric_number
```

```
                          1   2001 986491 1267750 84 0 407", header
= TRUE, sep = "")
FRA_2019_df <- read.table(text = "       Year petrol_number
diesel_gas_number full_mild_hybrid_number plugin_hybrid_number
battery_electric_number
                          1   2019 1283191 768825 95273 18442
42829", header = TRUE, sep = "")

GBR_2001_df <- read.table(text = "       Year petrol_number
diesel_gas_number full_mild_hybrid_number plugin_hybrid_number
battery_electric_number
                          1   2001 2019996 437628 719 0 29", header
= TRUE, sep = "")
GBR_2019_df <- read.table(text = "       Year petrol_number
diesel_gas_number full_mild_hybrid_number plugin_hybrid_number
battery_electric_number
                          1   2001 1467637 615619 155276 34758
37850", header = TRUE, sep = "")

# Converting wide data to long
DEU_2001_df <- pivot_longer(DEU_2001_df, -'Year', names_to =
'engine_type', values_to= 'count')
DEU_2019_df <- pivot_longer(DEU_2019_df, -'Year', names_to =
'engine_type', values_to= 'count')

FRA_2001_df <- pivot_longer(FRA_2001_df, -'Year', names_to =
'engine_type', values_to= 'count')
FRA_2019_df <- pivot_longer(FRA_2019_df, -'Year', names_to =
'engine_type', values_to= 'count')

GBR_2001_df <- pivot_longer(GBR_2001_df, -'Year', names_to =
'engine_type', values_to= 'count')
GBR_2019_df <- pivot_longer(GBR_2019_df, -'Year', names_to =
'engine_type', values_to= 'count')
```

Phase 4: Analyze

To identify trends in the different new engine types registered in the 6 European countries, the engine count data was plotted against year for each country, via R. A line graph was chosen for this purpose to clearly display the changes in new registration count over time.

New diesel engine registration analysis

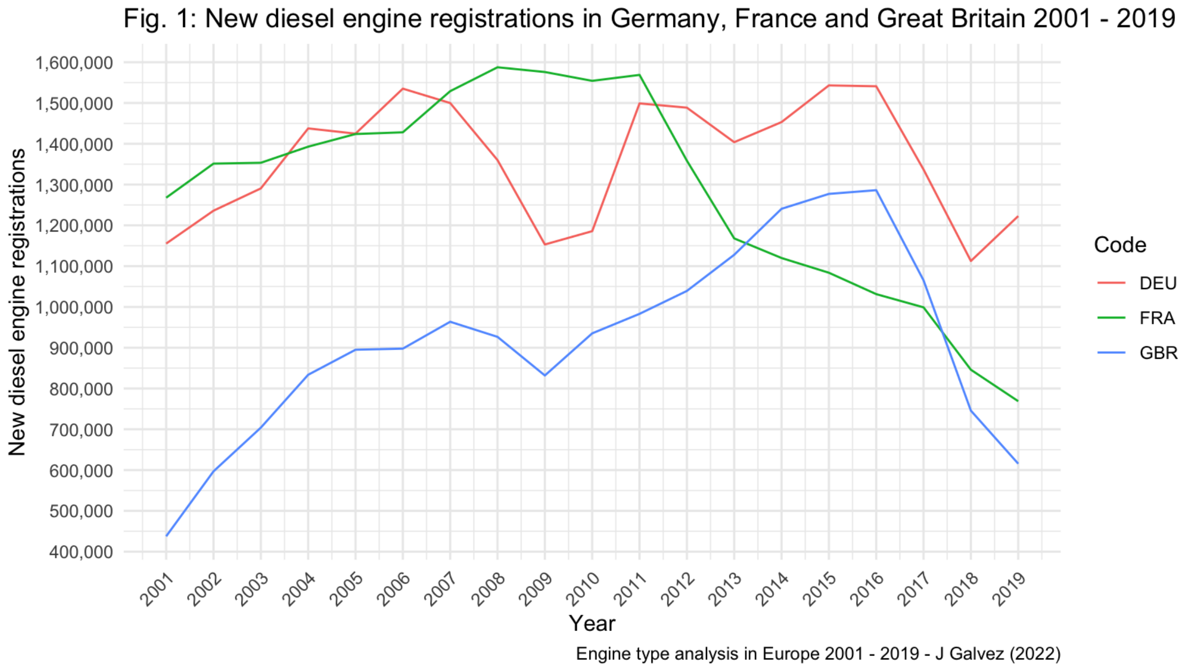
Observing figure 1, it is interesting to note that both Germany and Great Britain exhibit near homogeneous trends in new diesel engine registrations. Both Germany and Great Britain encountered a significant decrease from 2006 and 2007, respectively, until 2009 where they both experienced and increase again. During this time, Germany witnessed a 24.89% decrease in new diesel engine registrations, whereas Great Britain only witnessed a 13.67% decrease. This observation may be attributed to the great recession

[4] of 2007 to 2009 (N.B. the beginning and end years of the great recession varies amongst different nations, due to differing economies). In contrast, France impressively displayed an opposite trend during the same time period, showing a notable 11.16% increase.

Following the dip in 2009, Germany and Great Britain experienced a steady increase, both peaking in 2016 at 1,540,989.035 and 1,286,239 new registrations, respectively, before rapidly decreasing again. This time the decline is visibly greater, Germany experienced a 20.65% decrease, whereas Great Britain saw a staggering 52.15% decrease. France once again displays a unique trend, peaking earlier in 2011 at 1,568,874 new registrations and had decreased ever since, down to 768,825 new registrations in 2019 - a 51.00% decrease.

In September 2015, the German car manufacturer Volkswagen was investigated and later admitted to falsifying emissions tests within the US, dubbed as the 'diesel dupe' [5]. This sent ripples around the global automotive industry. As a result, 8.5 million Volkswagen units were recalled within Europe alone. Other car manufacturers such as Ford, BMW and Renault-Nissan group, denied using the same falsifying technology as Volkswagen. However, the magnitude of this scandal and the future implications of the investigative findings may be the reason new diesel engine registrations have been at a decline for Germany and Great Britain since 2016.

```
# Engine line graph plots for top 3 EU countries by GDP
# diesel engine count vs year
options(repr.plot.width = 12, repr.plot.height = 7)
diesel_top_3 <- ggplot(data = top_3_gdp_area, aes(x = Year, y =
diesel_gas_number, colour = Code)) +
  geom_line() +
  theme_minimal() +
  scale_x_continuous(n.breaks = 30) +
  scale_y_continuous(n.breaks = 10, labels = scales::comma) +
  labs(title = 'Fig. 1: New diesel engine registrations in Germany,
France and Great Britain 2001 - 2019',
       y = 'New diesel engine registrations',
       caption = 'Engine type analysis in Europe 2001 - 2019 - J
Galvez (2022)') +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
diesel_top_3
```



New petrol engine registration analysis

Looking at figure 2, attention is quickly drawn towards Germany's spike in new petrol engine registrations in 2009. This observation is especially shocking considering the entire world had just emerged from the great recession. That same year, it was reported that Sigmar Gabriel, the (at-the-time) German Federal Minister for the Environment, Nature Conservation and Nuclear Safety, cancelled the plans to introduce E10 fuel in Germany [6]. E10 fuel is a type of petrol which contains up to 10% renewable bioethanol, and was founded on the goal of formulating petrol which emits less CO2 [7]. It is possible that the drastic 52.61% increase of new petrol engine registrations to 2,628,234 from 2008 to 2009, was a result of anticipation of the release of E10 petrol.

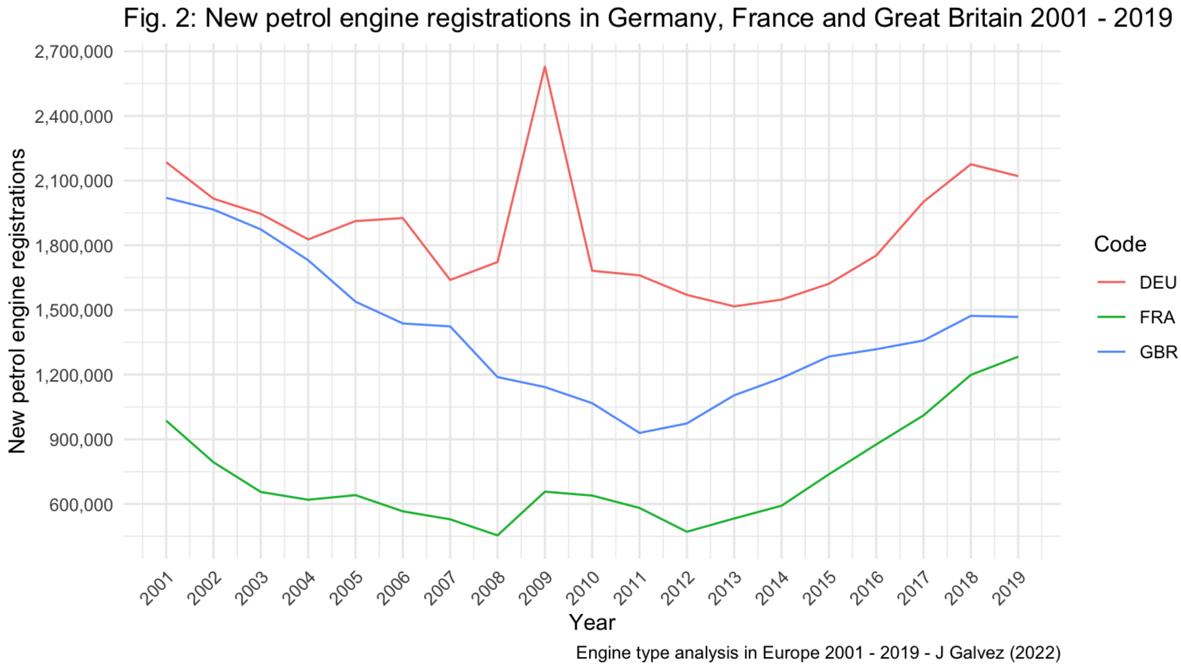
In comparison to France who did release E10 petrol in 2009 [8], the country only saw a 44.66% increase in new petrol engine registrations from 454,253.4431 in 2008 to 657,106.5859 in 2009. Still a significant increase, however dwarfed in actual new registration count compared to Germany. Nevertheless, in the 18 years between 2001 and 2019 France had a 30.08% increase of new petrol engine registrations, from 986,491 to 1,283,191.

Great Britain on the other hand showed a unpromising petrol engine market between 2001 to 2011. Steadily decreasing in new petrol engine registrations by 53.98%, from 2,019,996 to 929,590.4799. However, popularity for petrol engines grew again after 2011, gradually growing again to 1,467,637 in 2019 - a 257.88% increase, but still not meeting its former glory in 2001. Overall, between 2001 and 2019, Great Britain saw a drop in new petrol engine registration of 27.34%.

It's interesting to note that Great Britain and France exhibit similar trajectories. Albeit Great Britain having a dramatically steeper decline from 2011 to 2011, and France having a greater incline from 2012 to 2019.

```
# petrol engine count vs year
petrol_top_3 <- ggplot(data = top_3_gdp_area, aes(x = Year, y =
```

```
petrol_number, colour = Code)) +
  geom_line() +
  theme_minimal() +
  scale_x_continuous(n.breaks = 30) +
  scale_y_continuous(n.breaks = 10, labels = scales::comma) +
  labs(title = 'Fig. 2: New petrol engine registrations in Germany,
France and Great Britain 2001 - 2019',
       y = 'New petrol engine registrations',
       caption = 'Engine type analysis in Europe 2001 - 2019 - J
Galvez (2022)') +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
petrol_top_3
```



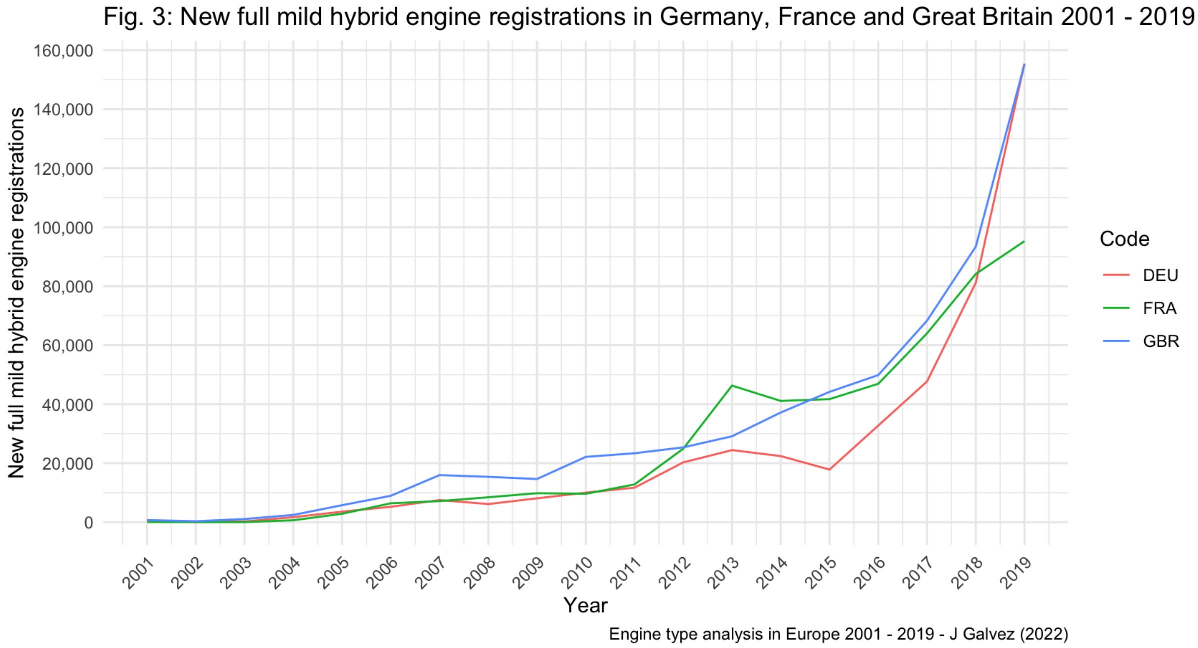
New full-mild hybrid engine registration analysis

Unlike new diesel and petrol engine registrations in France, Germany and Great Britain, full-mild hybrid engines show clear from fig. 3 a upward trend over time with little to no indications of a decline. The first full-mild hybrid car to be released on the market was Toyota's 1997 Prius, since then hybrid technology has come leaps and bounds to the machinery we know today. From 2001 to 2019, France observed a staggering 113,520.24% increase in new full-mild hybrid engine registrations, Germany a 25,753.76% increase and Great Britain a 21,696.11% increase.

Despite the rapid popularity and growth in new full-mild hybrid engine registrations in the 3 largest economies in Europe, it is important to note that the overall new registrations still pale in comparison to new diesel and petrol engine registrations.

```
# full_mild_hybrid_number engine count vs year
full_mild_top_3 <-  ggplot(data = top_3_gdp_area, aes(x = Year, y =
full_mild_hybrid_number, colour = Code)) +
  geom_line() +
  theme_minimal() +
```

```
  scale_x_continuous(n.breaks = 30) +
  scale_y_continuous(n.breaks = 10, labels = scales::comma) +
  labs(title = 'Fig. 3: New full mild hybrid engine registrations in
Germany, France and Great Britain 2001 - 2019',
       y = 'New full mild hybrid engine registrations',
       caption = 'Engine type analysis in Europe 2001 - 2019 - J
Galvez (2022)') +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
full_mild_top_3
```



New plug-in hybrid engine registration analysis

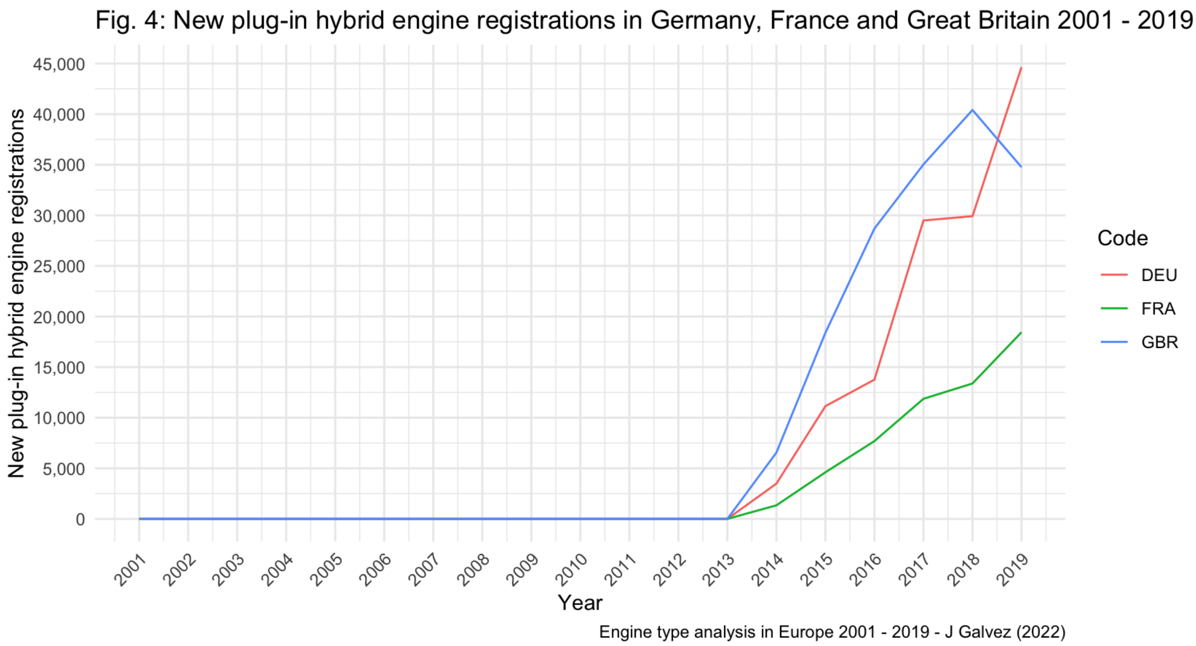
Despite the fact that plug-in hybrid engine vehicles and technology predates 2014, the obtained data regrettably contains no available data for their new registrations in France, Germany and Great Britain before then.

Similar to that of full-mild hybrid registrations, plug-in hybrids only display a clear upward trend between 2013 to 2019 for all three nations. France showed the most gradual growth amongst the three, with a 1,476.27% increase from 1,340 in 2014 to 18,442 in 2019. Great Britain experienced a 630% increase from 6,550 in 2014 to 34,758 in 2019. Lastly, Germany exhibited a 1,381.65% from 3,483 in 2014 to 44,640 in 2019.

Although France had a greater percentage growth than Great Britain, it is important to note that Great Britain had more new registrations in 2014 (the first year of new registrations for the three countries) than France. Great Britain still holds more overall new registrations that France in all years from 2014 to 2019.

```
# plugin_hybrid_number engine count vs year
plugin_top_3 <- ggplot(data = top_3_gdp_area, aes(x = Year, y =
plugin_hybrid_number, colour = Code)) +
  geom_line() +
```

```
  theme_minimal() +
  scale_x_continuous(n.breaks = 30) +
  scale_y_continuous(n.breaks = 10, labels = scales::comma) +
  labs(title = 'Fig. 4: New plug-in hybrid engine registrations in
Germany, France and Great Britain 2001 - 2019',
       y = 'New plug-in hybrid engine registrations',
       caption = 'Engine type analysis in Europe 2001 - 2019 - J
Galvez (2022)') +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
plugin_top_3
```



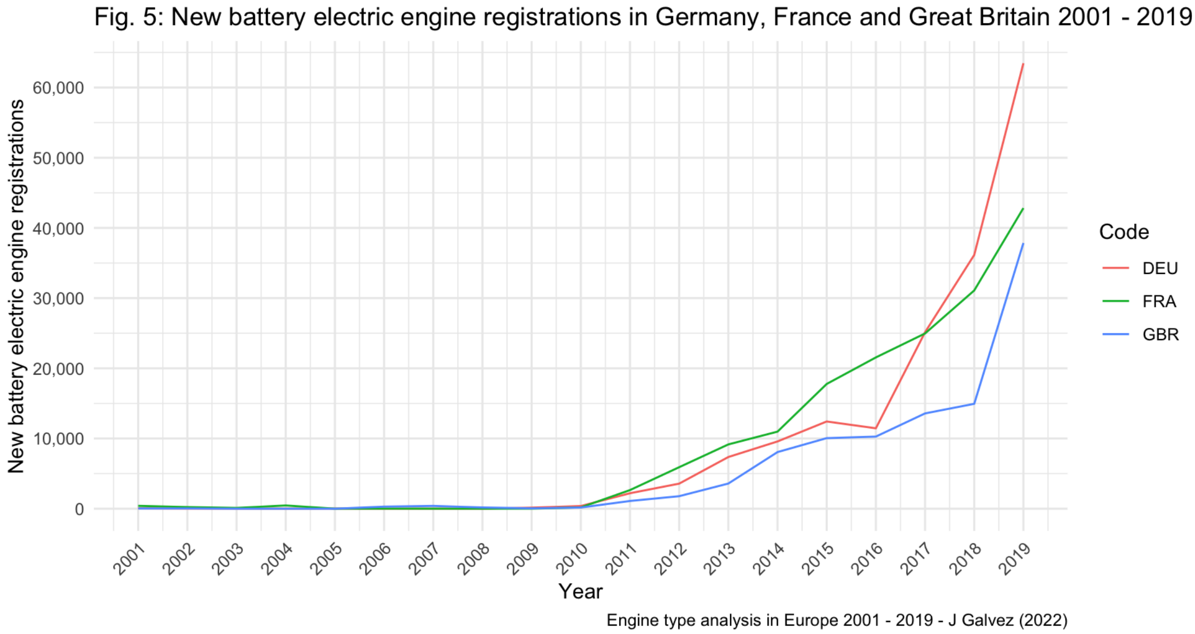
New battery electric engine registrations analysis

New battery electric engine registrations also display an overall positive trajectory over time within France, Germany and Great Britain from 2001 to 2019. Like their plug-in hybrid engine cousins, battery electric engines have long existed before 2001. In 1996, General Motors released the first mass-produced battery electric engine vehicle, the GM EV1 [9].

Between 2001 and 2010, all three countries had similar new registrations every year, it was not until 2011 where yearly new registrations began to vary among them. In 2010, France had 184 registrations, Germany had 380 and Great Britain had 167, a relatively even playing field. Germany arguably experienced the most noticeable dip in registrations in 2016, of 11,457 (down from 12,428 in 2015 - a 7.81% decrease). However, Germany also clearly had the greatest overall growth from 2010, shooting up to 63,461 in 2019 - an incredible 16,800.26% increase in 9 years. In the same time period, France also experienced a more impressive 23,376.63% increase (42,829 registrations in 2019) and Great Britain a grand 22,764.67% increase (37,850 registrations in 2019).

```
# battery_electric_number engine count vs year
battery_top_3 <- ggplot(data = top_3_gdp_area, aes(x = Year, y =
```

```
battery_electric_number, colour = Code)) +
  geom_line() +
  theme_minimal() +
  scale_x_continuous(n.breaks = 30) +
  scale_y_continuous(n.breaks = 10, labels = scales::comma) +
  labs(title = 'Fig. 5: New battery electric engine registrations in
Germany, France and Great Britain 2001 - 2019',
       y = 'New battery electric engine registrations',
       caption = 'Engine type analysis in Europe 2001 - 2019 - J
Galvez (2022)') +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
battery_top_3
```



New engine registration compositions

To effectively visualize the composition each country's new engine registration by type, a tree map was created for the 2001 and 2019. This visualization was favoured over a traditional pie or donut chart, because such a graphic represents composition of a thing, by angle. Humans innately have difficulty estimating quantity by angles, as opposed to (surface) area. A tree map directly links a graphic's surface area to the quantity/data value of a variable, making it more suited to visualize composition.

```
# Installing and loading packages
install.packages('treemapify', repos = "http://cran.us.r-
project.org")

##
## The downloaded binary packages are in
##  /var/folders/s4/1njj99vx55bgp7h4b35gmwk40000gn/T//Rtmpdw6cEl/
downloaded_packages

library(treemapify)
install.packages('ggpubr', repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/s4/1njj99vx55bgp7h4b35gmwk40000gn/T//Rtmpdw6cEl/
downloaded_packages
```

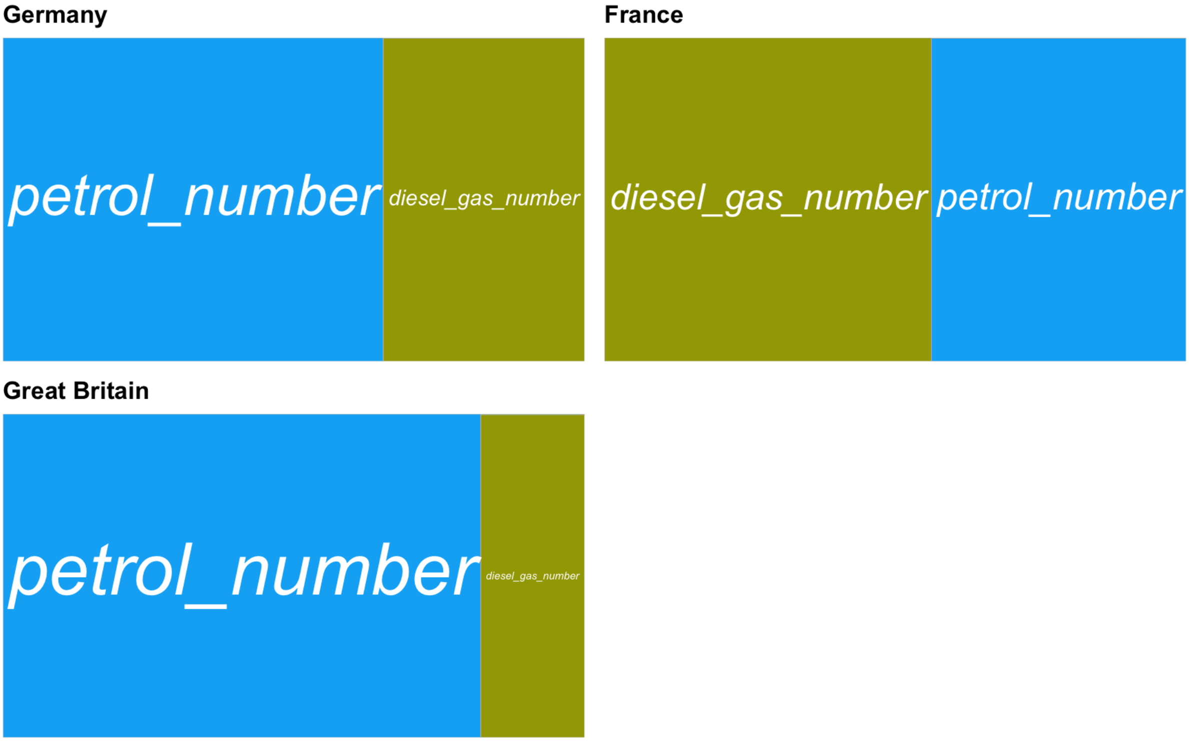
```
library(ggpubr)
```

```
##
## Attaching package: 'ggpubr'

## The following object is masked from 'package:cowplot':
##
##     get_legend
```

```
# Assigning individual tree maps
DEU_2001_tree <- ggplot(DEU_2001_df, aes(fill = engine_type, area =
count, label = engine_type)) +
  geom_treemap(show.legend = FALSE) +
  geom_treemap_text(fontface = "italic", colour = "white", place =
"centre",
                    grow = TRUE) +
  labs(title = 'Germany') +
theme(plot.title = element_text(face="bold"))

FRA_2001_tree <- ggplot(FRA_2001_df, aes(fill = engine_type, area =
count, label = engine_type)) +
  geom_treemap(show.legend = FALSE) +
  geom_treemap_text(fontface = "italic", colour = "white", place =
"centre",
                    grow = TRUE) +
  labs(title = 'France') +
theme(plot.title = element_text(face="bold"))

GBR_2001_tree <- ggplot(GBR_2001_df, aes(fill = engine_type, area =
count, label = engine_type)) +
  geom_treemap(show.legend = FALSE) +
  geom_treemap_text(fontface = "italic", colour = "white", place =
"centre",
                    grow = TRUE) +
  labs(title = 'Great Britain') +
theme(plot.title = element_text(face="bold"))

DEU_2019_tree <- ggplot(DEU_2019_df, aes(fill = engine_type, area =
count, label = engine_type)) +
  geom_treemap(show.legend = FALSE) +
  geom_treemap_text(fontface = "italic", colour = "white", place =
"centre",
                    grow = TRUE) +
  labs(title = 'Germany') +
theme(plot.title = element_text(face="bold"))
```

```
FRA_2019_tree <- ggplot(FRA_2019_df, aes(fill = engine_type, area =
count, label = engine_type)) +
  geom_treemap(show.legend = FALSE) +
  geom_treemap_text(fontface = "italic", colour = "white", place =
"centre",
                    grow = TRUE) +
  labs(title = 'France') +
theme(plot.title = element_text(face="bold"))

GBR_2019_tree <- ggplot(GBR_2019_df, aes(fill = engine_type, area =
count, label = engine_type)) +
  geom_treemap(show.legend = FALSE) +
  geom_treemap_text(fontface = "italic", colour = "white", place =
"centre",
                    grow = TRUE) +
  labs(title = 'Great Britain') +
theme(plot.title = element_text(face="bold"))

# Aggregating 3x 2001 tree maps into one graphic
tree_2001 <- ggarrange(DEU_2001_tree, FRA_2001_tree, GBR_2001_tree,
          labels = c("", "", ""),
          ncol = 2, nrow = 2)

# Aggregating 3x 2019 tree maps into one graphic
tree_2019 <- ggarrange(DEU_2019_tree, FRA_2019_tree, GBR_2019_tree,
          labels = c("", "", ""),
          ncol = 2, nrow = 2)

# Annotating aggregated 2001 tree maps
annotate_figure(tree_2001, top = text_grob("Fig. 6: Engine type
composition in 2001",
               color = "black", face = "bold", size = 14))
```

Fig. 6: Engine type composition in 2001




```r
# Annotating aggregated 2019 tree maps
annotate_figure(tree_2019, top = text_grob("Fig. 7: Engine type composition in 2019",
                color = "black", face = "bold", size = 14))
```

Fig. 7: Engine type composition in 2019

Observing both figures 6 and 7, it's made much more obvious the rise in popularity of full-mild hybrid, plug-in hybrid and battery electric engine vehicles have become in

2019, since 2001. New petrol and diesel engine registrations dominated the market in all three countries in 2001, with practically zero room for competitors. Germany and Great Britain had noticeably more new petrol engine registrations than diesel within their borders, but France's market resembled an approximate 60:40 new petrol to new diesel registration. By 2019, all three countries showed significant increases in new full-mild hybrid, plug-in hybrid and battery electric engine vehicle registrations. These tree maps compliment figures 1 - 5, further illustrating the favor for full-mild hybrid, plug-in hybrid and battery electric engine vehicles.

However, figures 6 and 7 display unique information which figures 1- 5 do not. Where the line graphs displays the trends for each engine type over the 19 year time frame, the tree maps capture a snapshot in time of each country's new vehicle engine registration market. Even though it is clear from figures 3 - 5, that full-mild hybrid, plug-in hybrid and battery electric engine vehicles are on the rise. Figures 6 and 7 illustrates to us clearly that they are still the minority in all three countries, in-terms of overall number of new vehicle engine registration per year.

Phase 5: Share and Act

Conclusions

Full analysis results can be found under 'Phase 4: Analyze', here's the TL;DR:

- **New diesel engine registrations:** Germany and Great Britain have similar trends and the 2008 recession caused a significant dip in 2009. Except for France, and actually peaked in 2011, but crashed ever since. Germany and Great Britain rose from the recession and peaked in 2016, only to crash ever since, likely following the 2015 'diesel dupe'.
- **New petrol engine registrations:** Despite the 2008 recession, Germany's spike in 2009 may have been in anticipation for E10 biofuel, but was eventually axed axed. Great Britain and France had a general decline from 2001 to 2011 and 2012, respectively, but had consistently been on the rise since then.
- **New full-mild hybrid, plug-in hybrid and battery electric engine vehicle registrations:** Each engine type has been on a dramatic increase in new registration since data had been available for them. In-terms of overall count, full-mild hybrids are dominant with battery-electric engines second and plug-in hybrids at last. Combined, they remain dwarfed in count by petrol and diesel engines, but show promise.

Recommendations and the Future

It is undeniable that the future is looking favored to the 'greener' options of full-mild hybrid, plug-in hybrid and battery electric engine vehicle. We are in an age of unprecedented global warming and environmental sustainability awareness. Prior to the knowledge we have today of greenhouse emissions, especially from transport, much of the world had no problem with using fossil fuels as a primary energy source. It is most likely the result of research, education and increased public understanding that traditional combustion engines (petrol and diesel) have seen an increase in competition from the aforementioned 'greener' alternative engines.

Coalitions such as COP27, the G20 and the EU have brought the topic of climate change further into the public, alongside environmental scientists and their research, with growing concerns each year. As a result of the contribution petrol and diesel engines have on climate change and global warming, the public and therefore the automobile market are likely only to see increased demands for greener options. This is reflected within the analysis portion of this case study.

It is critical to note, that this case study is not an environmental research paper, please refer to 'Phase 1: Ask' for the business tasks. As a result of this case study, it is advised to passenger automobile companies to grow and direct future endeavors towards their full-mild hybrid, plug-in hybrid and battery electric engine vehicles. Petrol and diesel engine vehicles are unlikely to disappear overnight, however, the data certainly points towards a lesser combustion engine future. Although, between the petrol and diesel it is suggested that petrol should take precedence in investment over diesel. Given the prospects of both combustion engines, diesel engines appear to head for a sharp decline, unable to recover from the diesel dupe. With such insights, it is worth it for automotive companies to pay close attention to current events, such as controversial investigations and/or development of 'greener' fuels. As such circumstances have, and will likely, have an impact on new engine registrations again.

Next steps

- Acquire data for 2020 - present
- Create a Tableau dashboard to illustrate my findings and potentially discover unrealized insights
- Produce a high-level/executive summary/presentation

References

[1] Number of registered passenger cars in Europe in 2018 and 2019, by country, Mathilde Carlier, 20 April 2022 (https://www.statista.com/statistics/452449/european-countries-number-of-registered-passenger-cars/)

[2] EU population up to nearly 513 million on 1 January 2018, Eurostat Press Office, 10 July 2018 (https://ec.europa.eu/eurostat/documents/2995521/9063738/3-10072018-BP-EN.pdf/ccdfc838-d909-4fd8-b3f9-db0d65ea457f#:~:text=On%201%20January%202018%2C%20the,million%20on%201%20January%202017.)

[3] missRanger: Fast Imputation of Missing Values, CRAN (https://cran.r-project.org/web/packages/missRanger/index.html#:~:text=missRanger%3A%20Fast%20Imputation%20of%20Missing,%2C%20D.J.%20and%20Buehlmann%2C%20P.)

[4] Transmission of the 2007–2008 financial crisis in advanced countries of the European Union, Kamila Tomcza, 09 March 2022 (https://onlinelibrary.wiley.com/doi/full/10.1111/boer.12338)

[5] Volkswagen: The scandal explained, BBC, 10 December 2015 (https://www.bbc.co.uk/news/business-34324772)

[6] Germany Cancels Plans to Introduce E10 in 2009, Rafael Seidl, 03 April 2008 (https://www.greencarcongress.com/2008/04/germany-cancels.html)

[7] E10 petrol explained, Department for Transport UK Gov., 25 February 2021, (https://www.gov.uk/guidance/e10-petrol-explained)

[8] Ethanol-based fuel becomes top-selling petrol in France, Sybille de La Hamaide, 07 November 2017 (https://www.reuters.com/article/us-france-ethanol/ethanol-based-fuel-becomes-top-selling-petrol-in-france-idUSKBN1D70HT)

[9] History of the electric car: from the first EV to the present day, Auto Express team, 4 May 2020 (https://www.autoexpress.co.uk/car-news/electric-cars/101002/history-of-the-ev-from-the-first-electric-car-to-the-present-day#:~:text=In%201996%2C%20General%20Motors%20released,of%20the%20industry’s%20key%20players)