Scenario

You are a junior data analyst working for a business intelligence consultant. You have been at your job for six months, and your boss feels you are ready for more responsibility. He has asked you to lead a project for a brand new client — this will involve everything from defining the business task all the way through presenting your data-driven recommendations. You will choose the topic, ask the right questions, identify a fresh dataset and ensure its integrity, conduct analysis, create compelling data visualizations, and prepare a presentation.

Ask

Five questions will guide your case study:

1. What type of company does your client represent, and what are they asking you to accomplish?
2. What are the key factors involved in the business task you are investigating?
3. What type of data will be appropriate for your analysis?
4. Where will you obtain that data?
5. Who is your audience, and what materials will help you present to them effectively?

Observation: I've made changes in the file using Excel, before importing to the Notebook

1. Name of columns to "year_XXXX"
2. Name of each Contry in the Country column
3. Numbers from "general" to "numbers"

STEP 1:

Installing packages

```
install.packages("tidyverse")
install.packages("ggplot2")

library(tidyverse)
library(ggplot2)
```

```
    Installing package into ‘/usr/local/lib/R/site-library’
    (as ‘lib’ is unspecified)

    Installing package into ‘/usr/local/lib/R/site-library’
    (as ‘lib’ is unspecified)
```

STEP 2:

Importing dataset

```
fish_df <- read.csv("/content/fish_catches2.csv")
head(fish_df)
```

A data.frame: 6 × 13

	Species	Area	Units	Country	year_2014	year_2013	year_2012	year_2011	year_2010	year_2009	year_2008	year_2007	year_2006
	<chr>	<chr>	<chr>	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	ANF	27	TLW	Belgium	993	1633	1716	1279	1031	853	964	1363	1193
2	ANF	27.4	TLW	Belgium	217	137	133	116	131	140	185	181	141
3	ANF	27.4.A	TLW	Belgium	0	0	0	0	0	0	0	0	0
4	ANF	27.4.B	TLW	Belgium	213	135	131	111	124	134	181	179	138
5	ANF	27.4.C	TLW	Belgium	4	2	2	6	7	6	4	3	3
6	ANF	27.7	TLW	Belgium	491	1124	1382	966	721	518	585	1040	907

STEP 3:

STEP 3:

transformation

```
fish_df <- fish_df %>%
  mutate(Country = recode(Country,
                  " UK (Channel Island Guernsey)" = "UK (Guernsey)" ,
                  " UK (Channel Island Jersey)" = "UK (Jersey)",
                  "LY" = "Libya",
                  "CN" = "China",
                  "JP" = "Japan" ,
                  "TW" = "Taiwan"))
```

i've transformed the names for shorter one's so the graphic can look better

```
unique_fish <- unique(fish_df$Species)
length(unique_fish)
```

```
909
```

Insight 1: 909 species on the data frame

Lets see the Countrys

```
unique_country <- unique(fish_df$Country)
print(unique_country)
length(unique_country)
```

```
 [1] " Belgium"          "China"             " Germany"
 [4] " Denmark"          " Estonia"          " Spain"
 [7] " Finland"          " Faroe Islands"    " France"
[10] "UK (Guernsey)"     " Greenland"        " Ireland"
[13] " UK (Isle of Man)" " Iceland"          "UK (Jersey)"
[16] "Japan"             " Lithuania"        " Latvia"
[19] "Libya"             " Netherlands"      " Norway"
[22] " Poland"           " Portugal"         " Russia"
[25] " Sweden"           "Taiwan"            " United Kingdom"
27
```

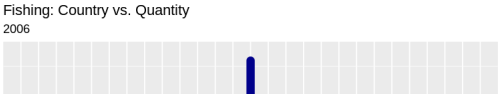
Insight 2: 27 countries in the data frame

STEP 4:

Analysis

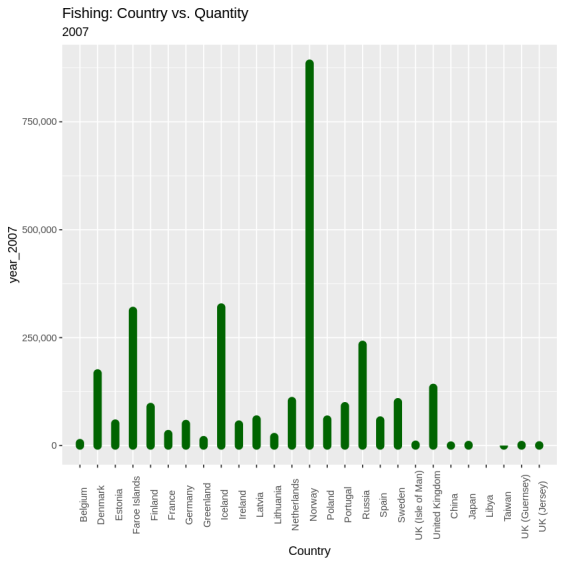
```
ggplot(data=fish_df, mapping = aes(x=Country, y=year_2006)) +
  geom_smooth(stat = "identity", color = "darkblue", position = position_dodge(width = 2), size=3.5) +
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::comma) +
  labs(title="Fishing: Country vs. Quantity", subtitle = "2006")
```

```
Warning message:
“`position_dodge()` requires non-overlapping x intervals”
```



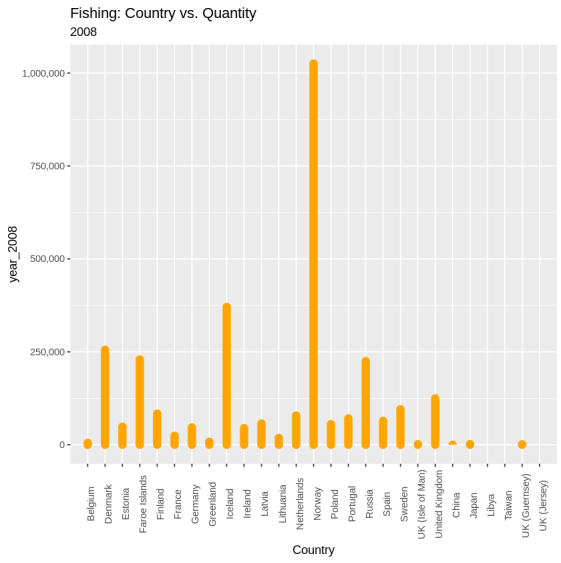

```
ggplot(data=fish_df, mapping = aes(x=Country, y=year_2007)) + geom_smooth(stat = "identity", color = "darkgreen", position = position_dodge(w
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::comma) +
  labs(title="Fishing: Country vs. Quantity", subtitle = "2007")
```

```
Warning message:
“`position_dodge()` requires non-overlapping x intervals”
```



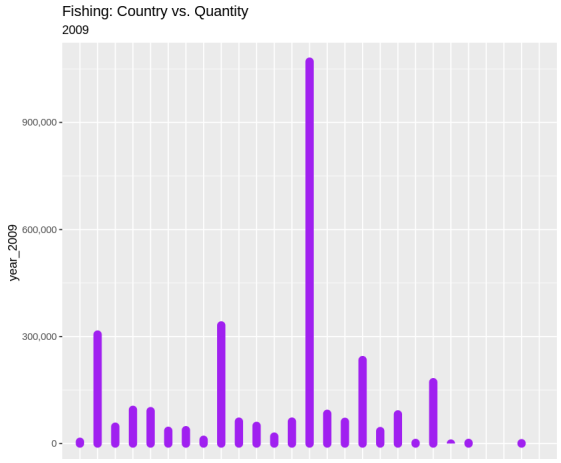

```
ggplot(data=fish_df, mapping = aes(x=Country, y=year_2008)) + geom_smooth(stat = "identity", color = "orange", position = position_dodge(widt
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::comma) +
  labs(title="Fishing: Country vs. Quantity", subtitle = "2008")
```

```
Warning message:
“`position_dodge()` requires non-overlapping x intervals”
```



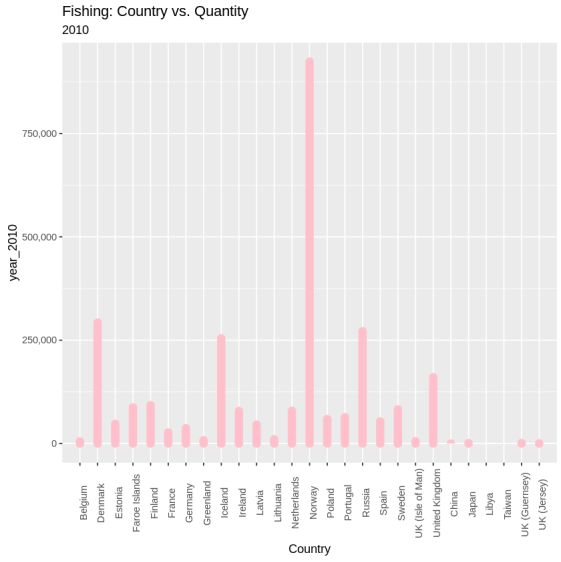

```
ggplot(data=fish_df, mapping = aes(x=Country, y=year_2009)) + geom_smooth(stat = "identity", color = "purple", position = position_dodge(widt
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::comma) +
  labs(title="Fishing: Country vs. Quantity", subtitle = "2009")
```

```
Warning message:
“`position_dodge()` requires non-overlapping x intervals”
```




```
ggplot(data=fish_df, mapping = aes(x=Country, y=year_2010)) + geom_smooth(stat = "identity", color = "pink", position = position_dodge(width
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::comma) +
  labs(title="Fishing: Country vs. Quantity", subtitle = "2010")
```

```
Warning message:
“`position_dodge()` requires non-overlapping x intervals”
```

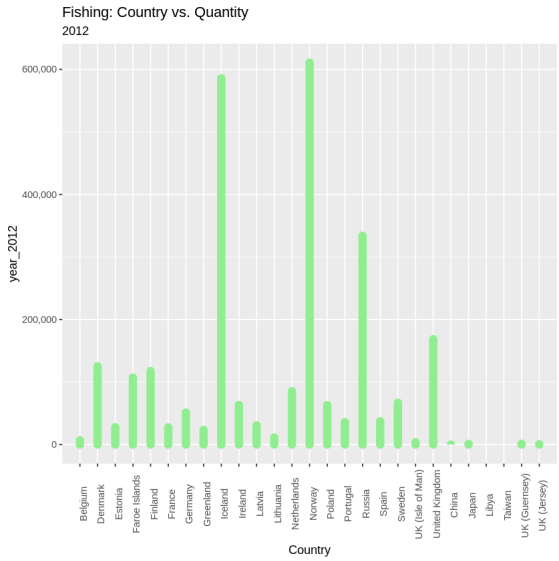



```
ggplot(data=fish_df, mapping = aes(x=Country, y=year_2011)) + geom_smooth(stat = "identity", color = "lightblue", position = position_dodge(w
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::comma) +
  labs(title="Fishing: Country vs. Quantity", subtitle = "2011")
```

```
Warning message:
“`position_dodge()` requires non-overlapping x intervals”
```

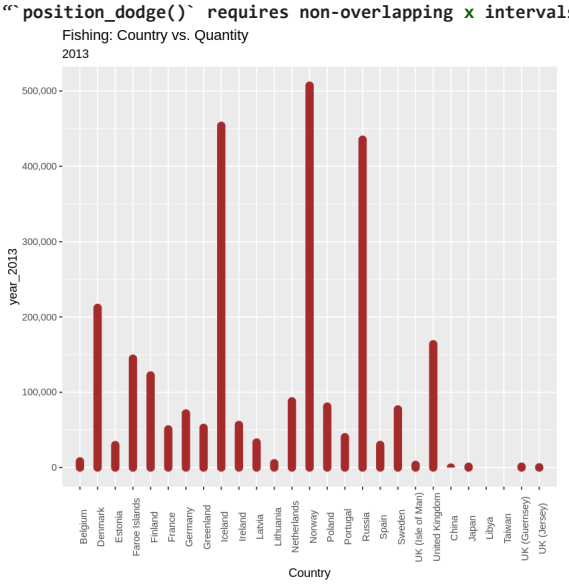

```
ggplot(data=fish_df, mapping = aes(x=Country, y=year_2012)) + geom_smooth(stat = "identity", color = "lightgreen", position = position_dodge(
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::comma) +
  labels(title="Fishing: Country vs. Quantity", subtitle = "2012")
```

```
Warning message:
“`position_dodge()` requires non-overlapping x intervals”
```



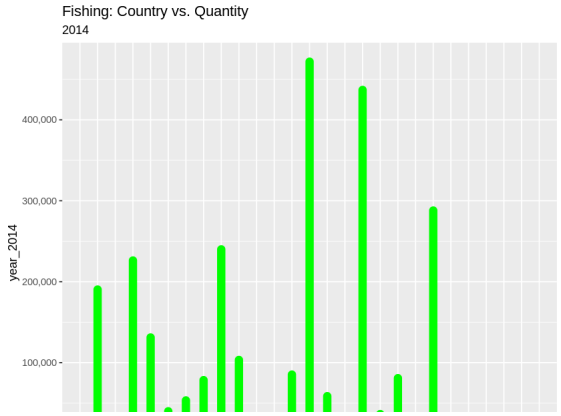

```
ggplot(data=fish_df, mapping = aes(x=Country, y=year_2013)) + geom_smooth(stat = "identity", color = "brown", position = position_dodge(width
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::comma) +
  labels(title="Fishing: Country vs. Quantity", subtitle = "2013")
```

```
Warning message:
“`position_dodge()` requires non-overlapping x intervals”
```




```
ggplot(data=fish_df, mapping = aes(x=Country, y=year_2014)) + geom_smooth(stat = "identity", color = "green", position = position_dodge(width
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::comma) +
  labels(title="Fishing: Country vs. Quantity", subtitle = "2014")
```

```
Warning message:
“`position_dodge()` requires non-overlapping x intervals”
```




```
fish_df %>%
  group_by(Country) %>%
  drop_na() %>%
  summarize(sum_2006 = sum(year_2006),
            sum_2007 = sum(year_2007),
            sum_2008 = sum(year_2008),
            sum_2009 = sum(year_2009),
            sum_2010 = sum(year_2010),
            sum_2011 = sum(year_2011),
            sum_2012 = sum(year_2012),
            sum_2013 = sum(year_2013),
            sum_2014 = sum(year_2014))
```

A tibble: 27 × 10

Country	sum_2006	sum_2007	sum_2008	sum_2009	sum_2010	sum_2011	sum_2012	sum_2013	sum_2014
<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>

Insight 3:

1. 2006 to 2008 there was a increasing number of fishing, from 700,000 up to almost 1,500,000
2. 2008 to 2014 there was a decreasing number of fishing, from almost 1,500,000 to almost 600,000

Conclusion Insight 3: as there year passes by, there is a significant decreasing number of fishing

Insight 4:

1. Norway was the leader of fishing in all the years
2. UK (Channel Island Jersey), UK (Channel Island Guernsey), UK (Isle of Man), Belgium, Libya, Taiwan, China, Japan has the less fishing in all years

Ireland	698563	727428	672211	862293	971306	655522	900720	799250	889214
Latvia	335881	362052	349543	319252	305034	263505	236231	251297	246703
Lithuania	121772	177632	136062	150592	78157	85687	81503	55562	160215
Netherlands	1145406	1147681	986276	859384	942744	778568	976734	990876	903008
Norway	7987448	8424718	8608234	9055430	9369447	7797434	7587647	7019769	7751748
Poland	467510	493128	393994	535033	462437	461140	503474	563460	496597
Portugal	582531	662743	591149	510638	569835	564209	518447	519942	430327
Russia	3441168	3258151	3164049	3393475	3645490	3421594	3485404	3579782	3682516
Spain	1163588	1025846	1203310	1062400	1195609	1136200	899403	1019984	1125042
Sweden	967208	871512	854296	752790	770899	652482	547125	632820	622133
UK (Isle of Man)	3627	11280	8310	11208	24028	20013	19558	19039	13369
United Kingdom	1868627	1869519	1713462	1768320	1822181	1794666	1891320	1889970	2315921
China	274	194	236	82	84	74	72	76	0
Japan	3936	5440	5658	4744	2996	2298	2602	2438	0
Libya	98	0	0	0	0	0	0	0	0

✓ 0 s completado a las 12:09