

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.2
```

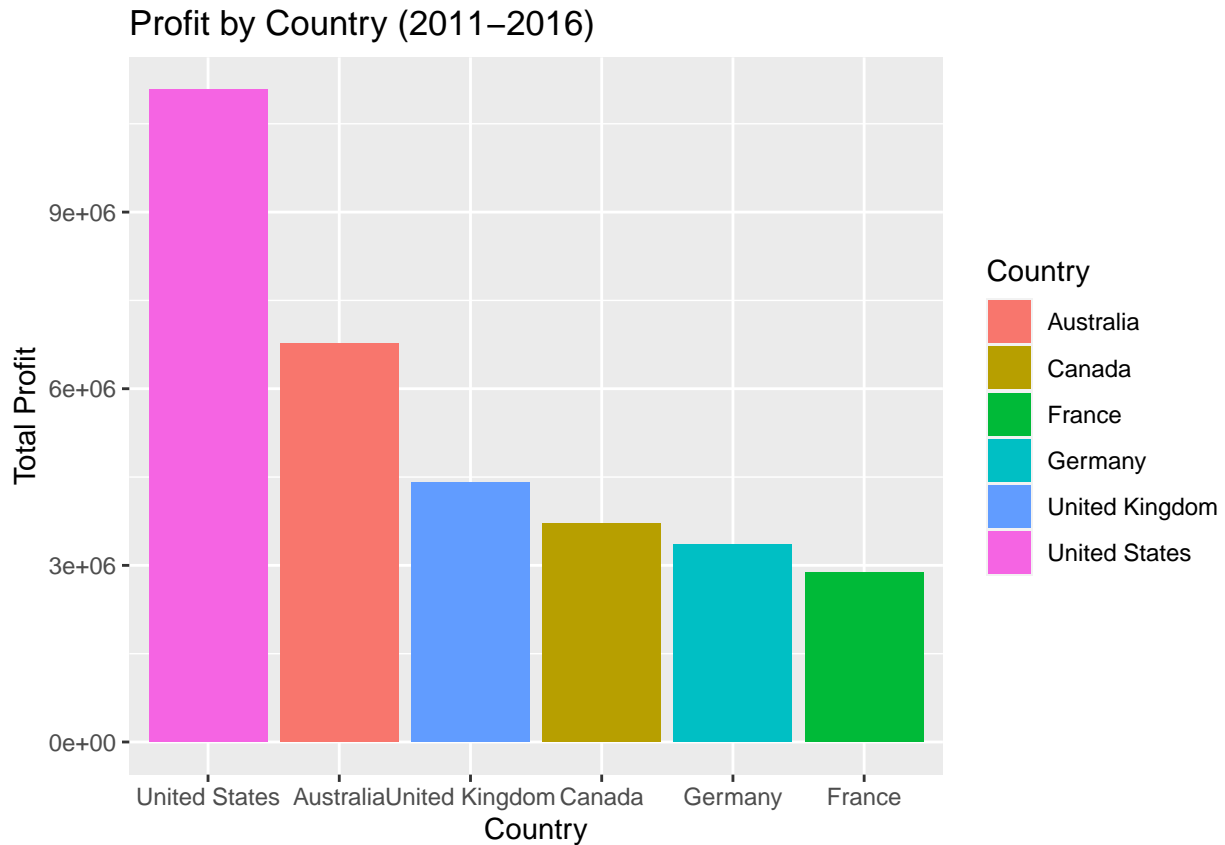
```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
sales <- read.csv("/cloud/project/Sales.csv")
head(sales)
```

```
##      Date Day   Month Year Customer_Age   Age_Group Customer_Gender
## 1 2013-11-26 26 November 2013          19   Youth (<25)             M
## 2 2015-11-26 26 November 2015          19   Youth (<25)             M
## 3 2014-03-23 23   March 2014          49 Adults (35-64)             M
## 4 2016-03-23 23   March 2016          49 Adults (35-64)             M
## 5 2014-05-15 15     May 2014          47 Adults (35-64)             F
## 6 2016-05-15 15     May 2016          47 Adults (35-64)             F
##      Country      State Product_Category Sub_Category      Product
## 1   Canada British Columbia   Accessories   Bike Racks Hitch Rack - 4-Bike
## 2   Canada British Columbia   Accessories   Bike Racks Hitch Rack - 4-Bike
## 3 Australia New South Wales   Accessories   Bike Racks Hitch Rack - 4-Bike
## 4 Australia New South Wales   Accessories   Bike Racks Hitch Rack - 4-Bike
## 5 Australia New South Wales   Accessories   Bike Racks Hitch Rack - 4-Bike
## 6 Australia New South Wales   Accessories   Bike Racks Hitch Rack - 4-Bike
##      Order_Quantity Unit_Cost Unit_Price Profit Cost Revenue
## 1                8         45        120    590 360    950
## 2                8         45        120    590 360    950
## 3               23         45        120   1366 1035   2401
## 4               20         45        120   1188  900   2088
## 5                4         45        120    238  180    418
## 6                5         45        120    297  225    522
```

Data set used: <https://www.kaggle.com/datasets/sadiqshah/bike-sales-in-europe>

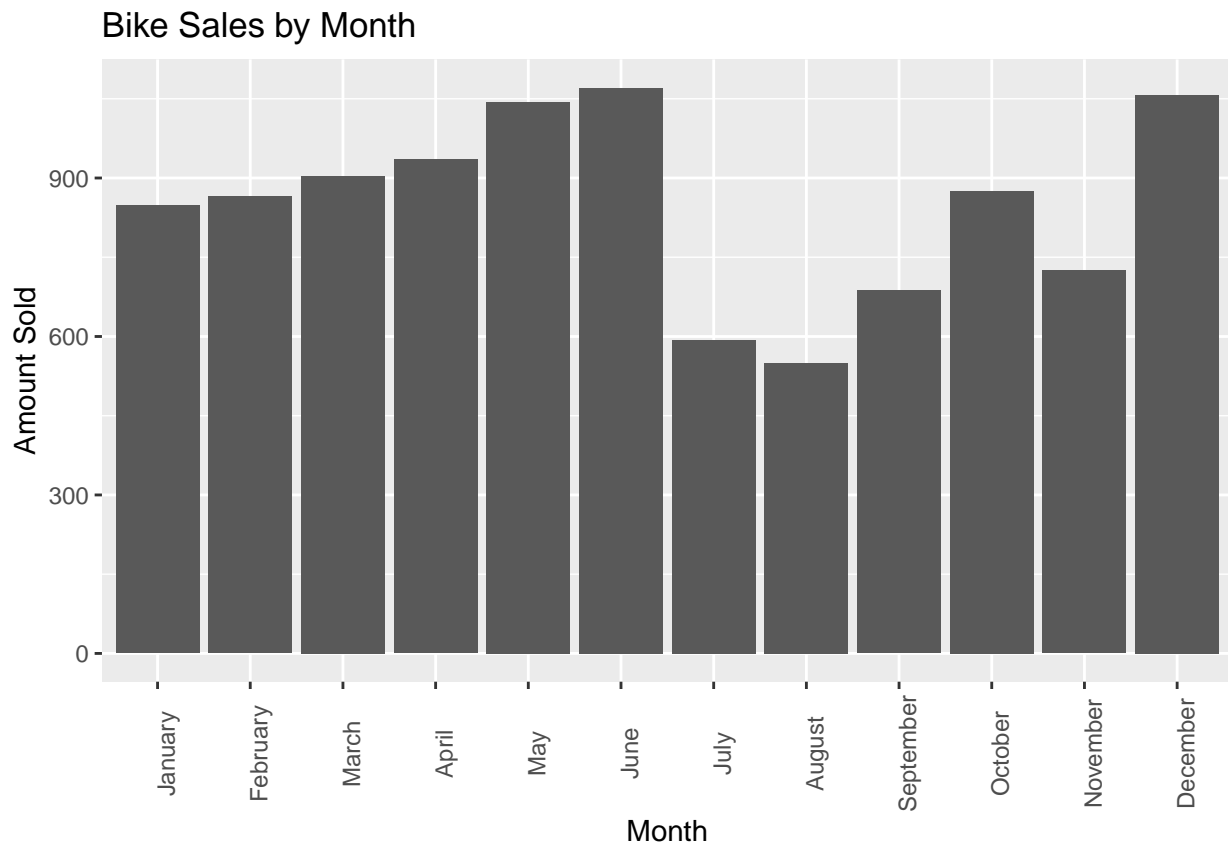
```
sales %>%
  group_by(Country) %>%
  summarise(total_profit = sum(Profit)) %>%
  ggplot(., aes(x = reorder(Country, -total_profit), y = total_profit, fill = Country))+
  geom_col()+
  xlab("Country")+
  ylab("Total Profit")+
  ggtitle("Profit by Country (2011-2016)")
```



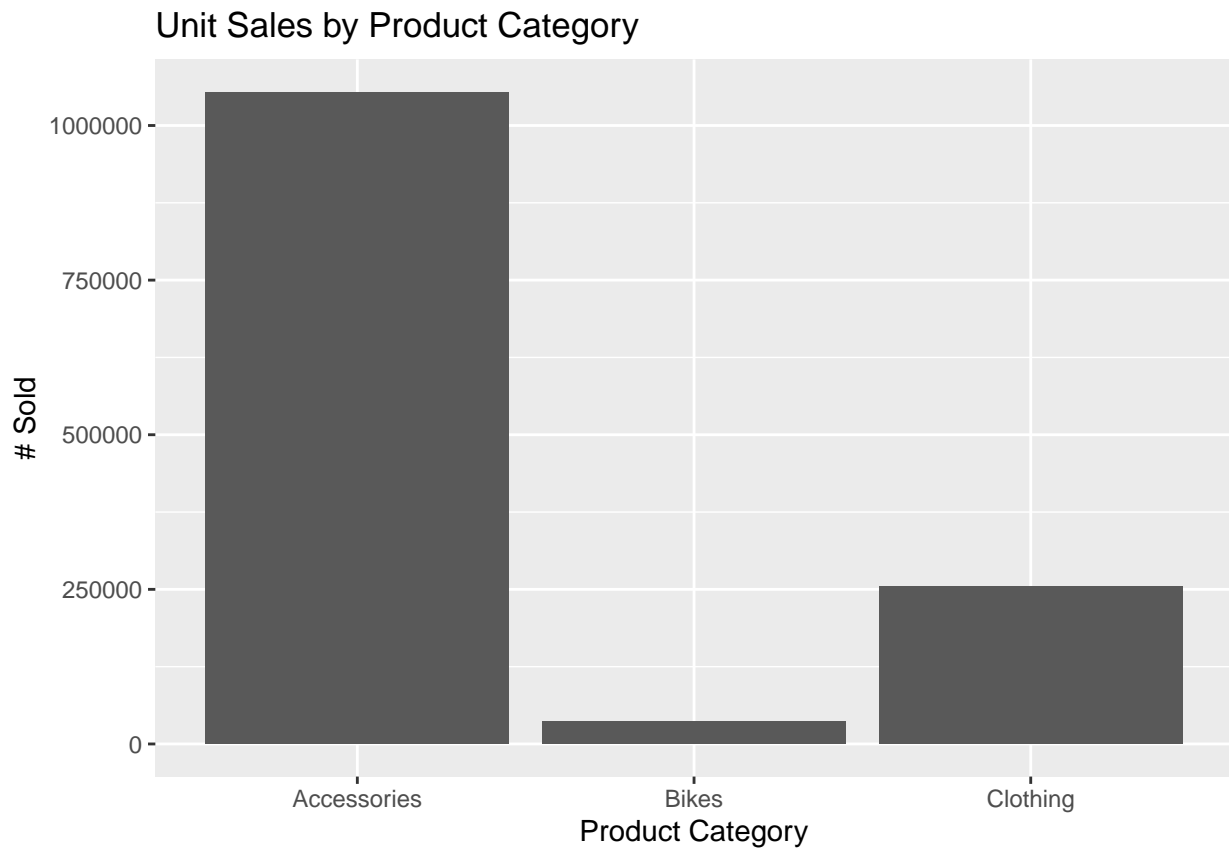
We were interested in which country generated the most profit from bike sales, so we graphed the grouped

```
sales$Month <- factor(sales$Month,
                      levels = c("January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December"))

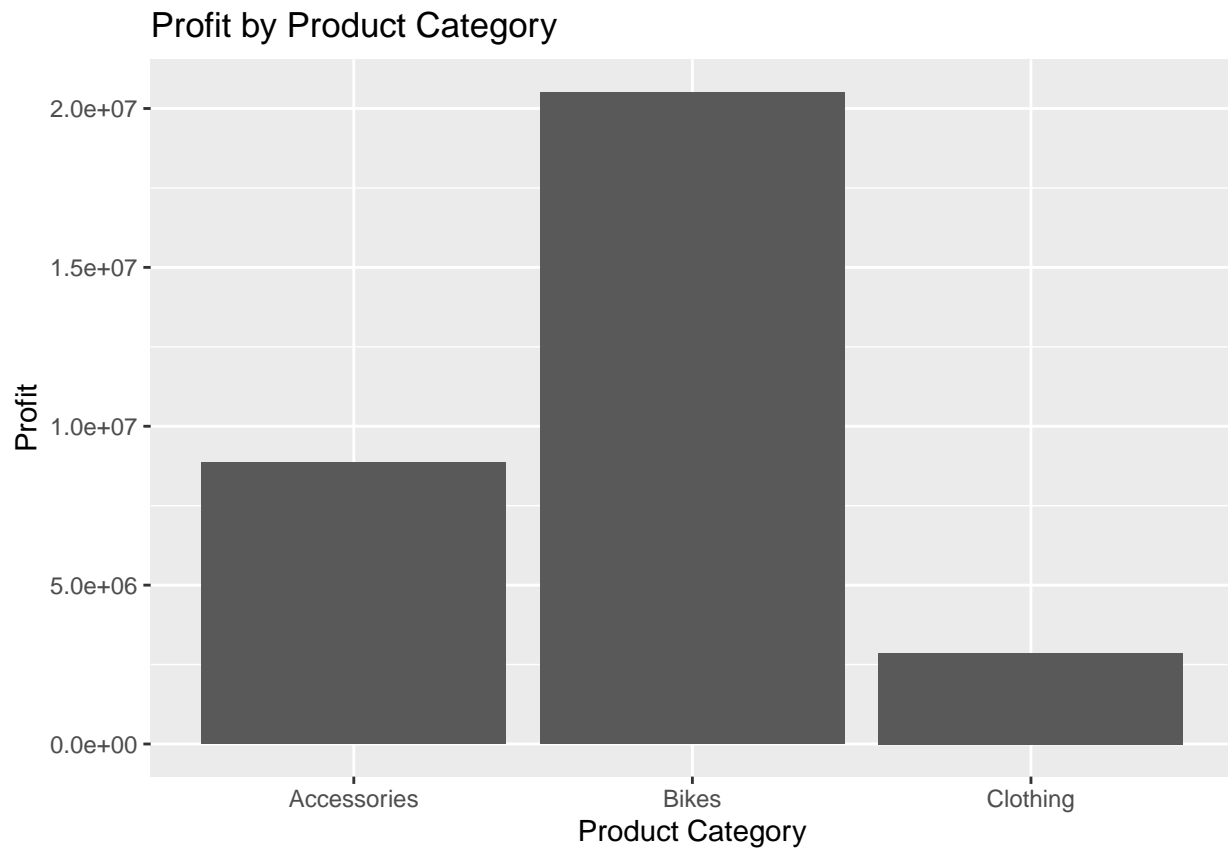
sales %>%
  arrange(desc(Month)) %>%
  filter(Product_Category == "Bikes") %>%
  group_by(Month) %>%
  filter(Country == "Australia") %>%
  summarise(n = sum(Order_Quantity)) %>%
  ggplot(., aes(Month, n)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 90)) +
  ggtitle("Bike Sales by Month")+
  ylab("Amount Sold")+
  xlab("Month")
```



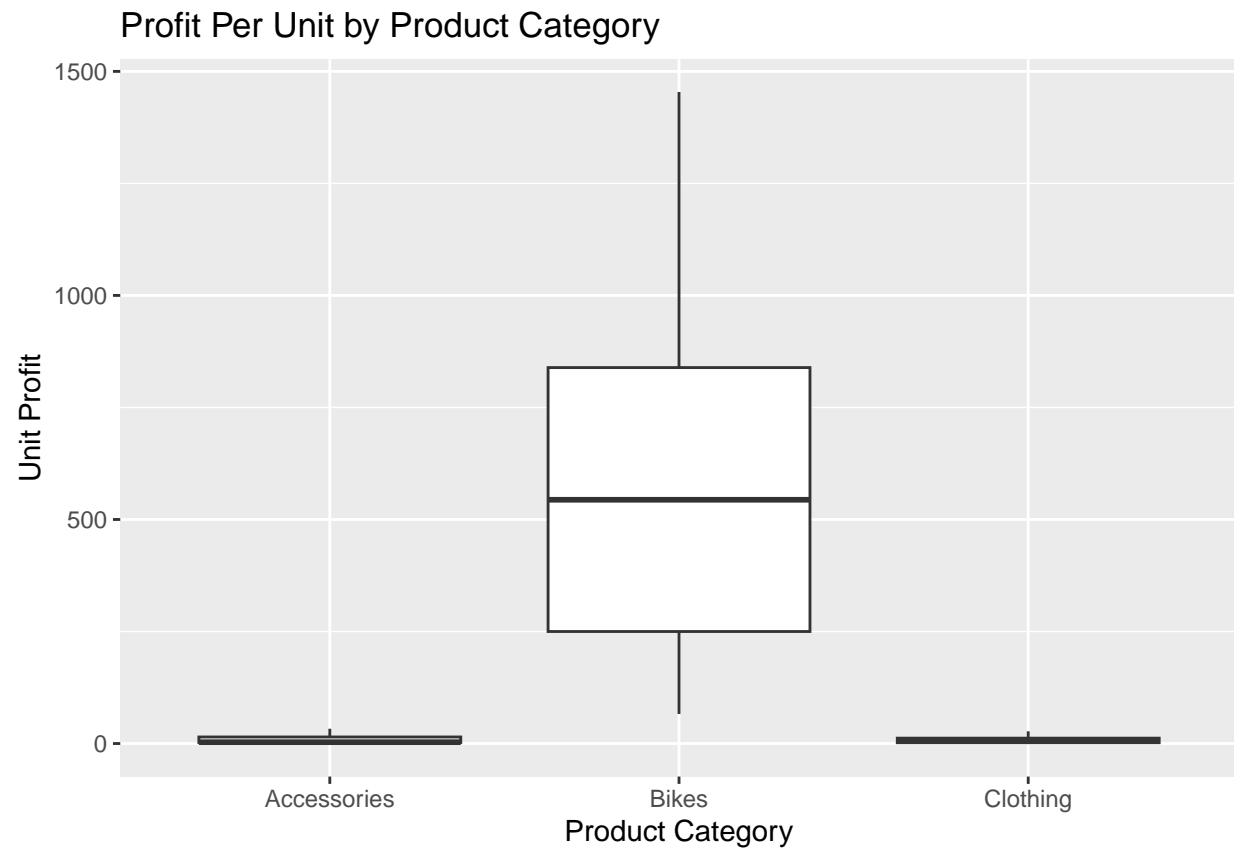
```
amounts_per_category <- sales %>%  
  group_by(Product_Category) %>%  
  summarise(amount_sold = sum(Order_Quantity))  
amounts_per_category <- as.data.frame(amounts_per_category)  
  
ggplot(amounts_per_category, aes(x = Product_Category, y = amount_sold)) +  
  geom_col() +  
  ylab("# Sold") +  
  xlab("Product Category") +  
  ggtitle("Unit Sales by Product Category")
```



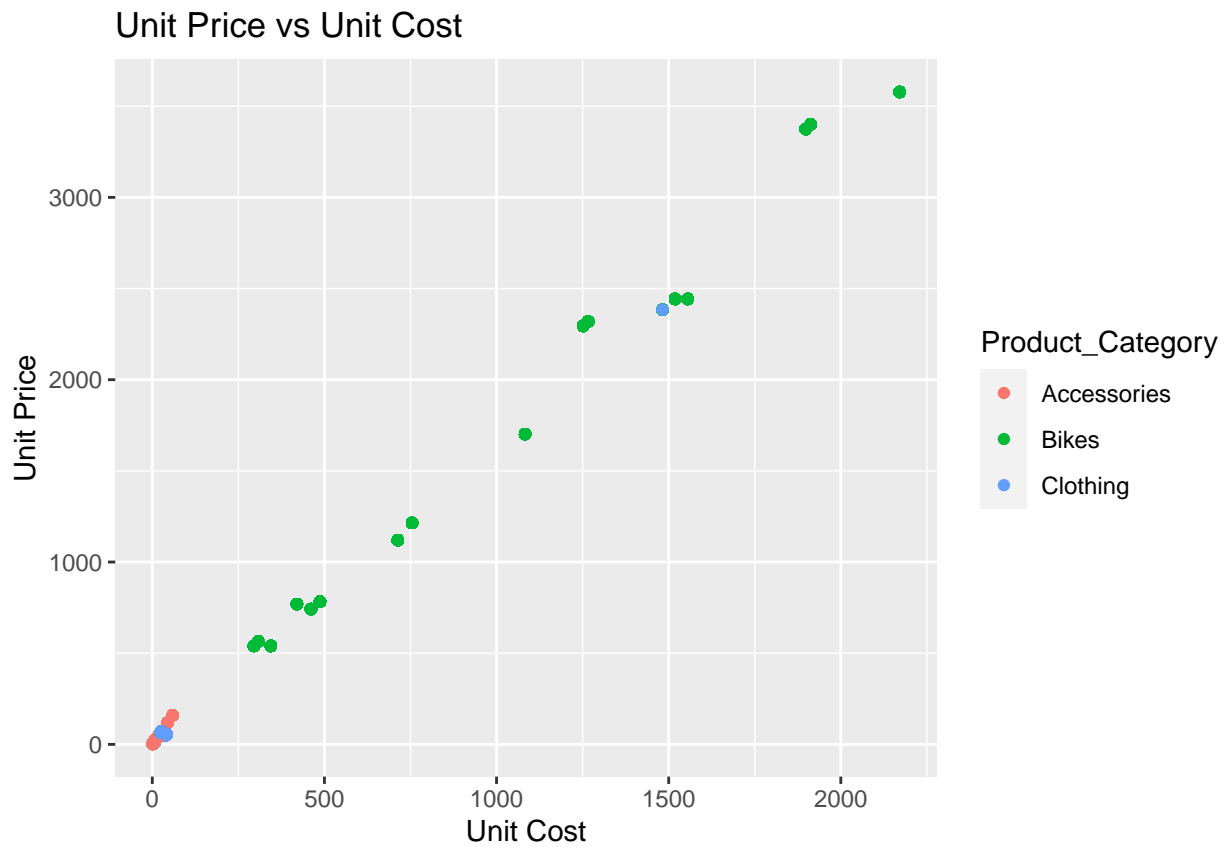
```
ggplot(sales, aes(x = Product_Category, y = Profit)) +  
  geom_col() +  
  ylab("Profit") +  
  xlab("Product Category") +  
  ggtitle("Profit by Product Category")
```



```
ggplot(sales, aes(x = Product_Category, y = Profit / Order_Quantity)) +  
  geom_boxplot(outlier.alpha = FALSE) +  
  ylab("Unit Profit") +  
  xlab("Product Category") +  
  ggtitle("Profit Per Unit by Product Category")
```



```
ggplot(sales, aes(x = Unit_Cost, y = Unit_Price, color = Product_Category)) +  
  geom_point() +  
  ylab("Unit Price") +  
  xlab("Unit Cost") +  
  ggtitle("Unit Price vs Unit Cost")
```



```
ggplot(sales, aes(x = Unit_Cost, y = Unit_Price, color = Product_Category)) +  
  geom_point() +  
  xlim(0, 60) +  
  ylim(0, 175) +  
  ylab("Unit Price") +  
  xlab("Unit Cost") +  
  ggtitle("Unit Price vs Unit Cost")
```

Warning: Removed 25984 rows containing missing values (`geom_point()`).

