



# TRỰC QUAN HÓA DỮ LIỆU NÔNG NGHIỆP VIỆT NAM

---

Môn: Trực quan hóa dữ liệu

# 1. THU THẬP DỮ LIỆU

Được thu thập từ tổng cục thống kê Việt Nam

Bao gồm dữ liệu về sản lượng và diện tích:

- Trồng cây lương thực (lúa, ngô, khoai, sắn).
- Chăn nuôi.
- Lâm nghiệp.
- Ngư nghiệp.



Logo tổng cục thống kê

## 2. MÔ TẢ TẬP DỮ LIỆU

Tập dữ liệu có 1775 dòng và 21 cột

Bao gồm dữ liệu về trồng lương thực (lúa, ngô khoai, sắn), chăn nuôi, lâm nghiệp và ngư nghiệp của một đơn vị địa lý (tỉnh, khu vực, cả nước) trong 1 năm trong khoảng 2000-2019.

	Năm	Tỉnh, thành phố	DT_khoai	SL_khoai	DT_thuysan
1770	Sơ bộ 2019	Cần Thơ	..	..	7.1
1771	Sơ bộ 2019	Hậu Giang	0.1	1.4	7.8
1772	Sơ bộ 2019	Sóc Trăng	1.5	19.3	78.9
1773	Sơ bộ 2019	Bạc Liêu	0.5	3.5	140.5
1774	Sơ bộ 2019	Cà Mau	0.2	0.9	305.0

5 dòng cuối của dữ liệu

### 3. TRỰC QUAN VÀ PHÂN TÍCH DỮ LIỆU TRỒNG CÂY LƯƠNG THỰC

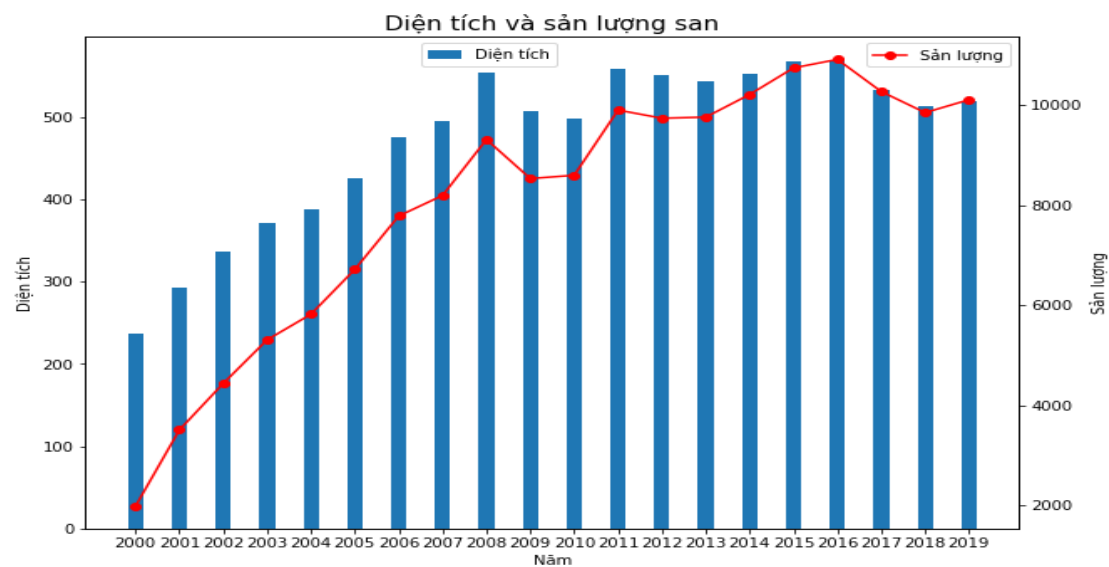
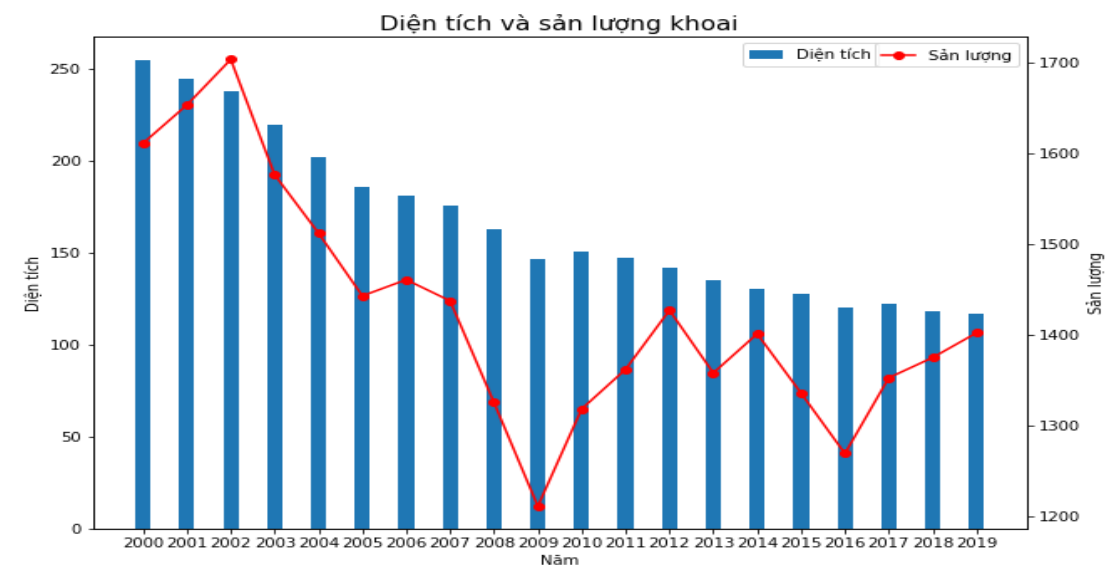
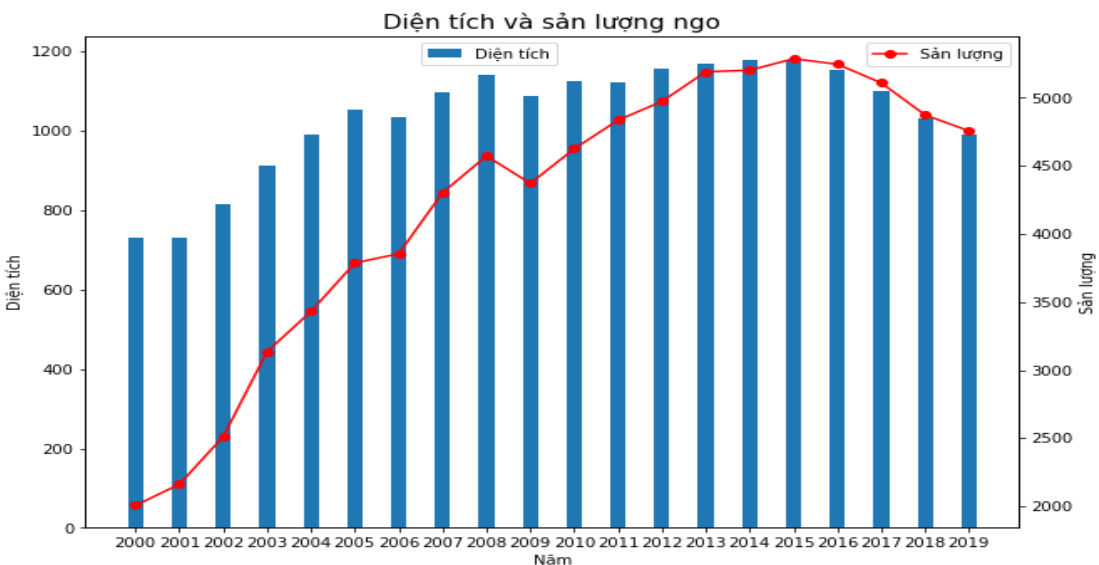
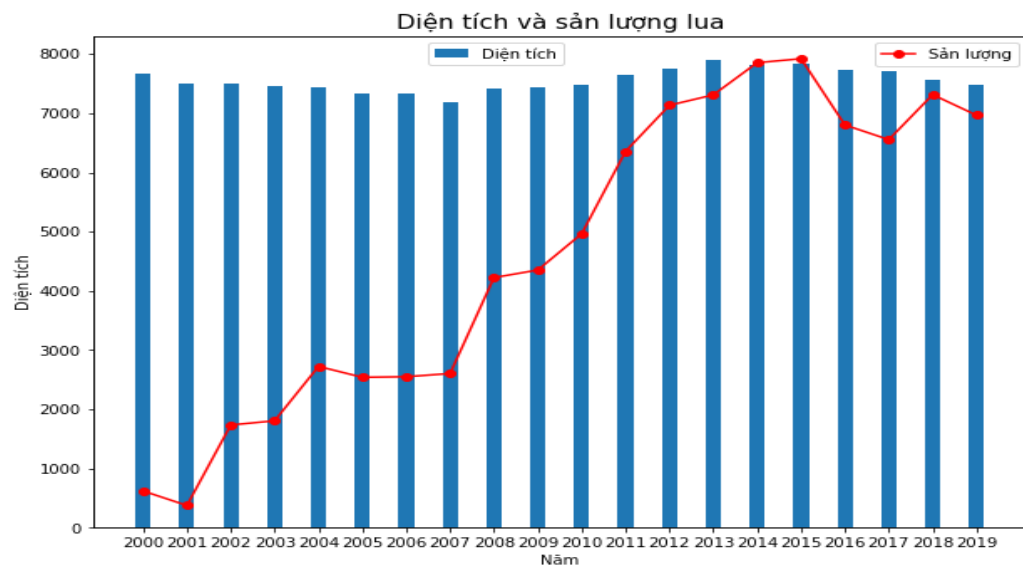
Dữ liệu trồng cây lương thực bao gồm:

Diện tích (nghìn ha)

Sản lượng (nghìn tấn)

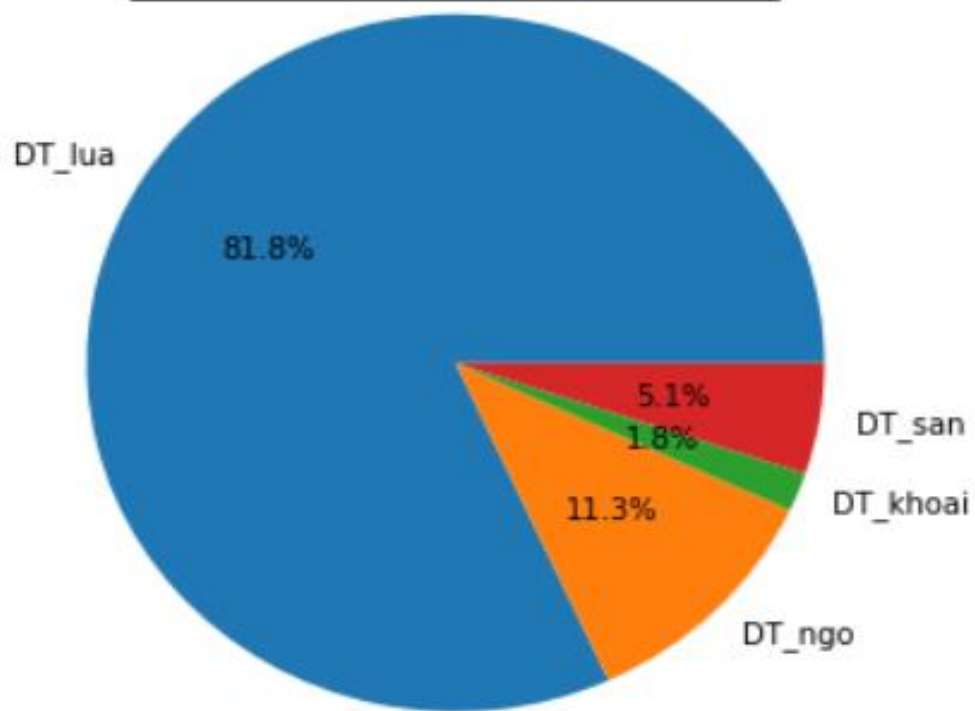
Của lúa, ngô, sắn, khoai trên một đơn vị diện tích lãnh thổ trong 1 năm trong khoảng 2000-2019

# Sản lượng và diện tích lúa về cây lương thực qua các năm trên cả nước

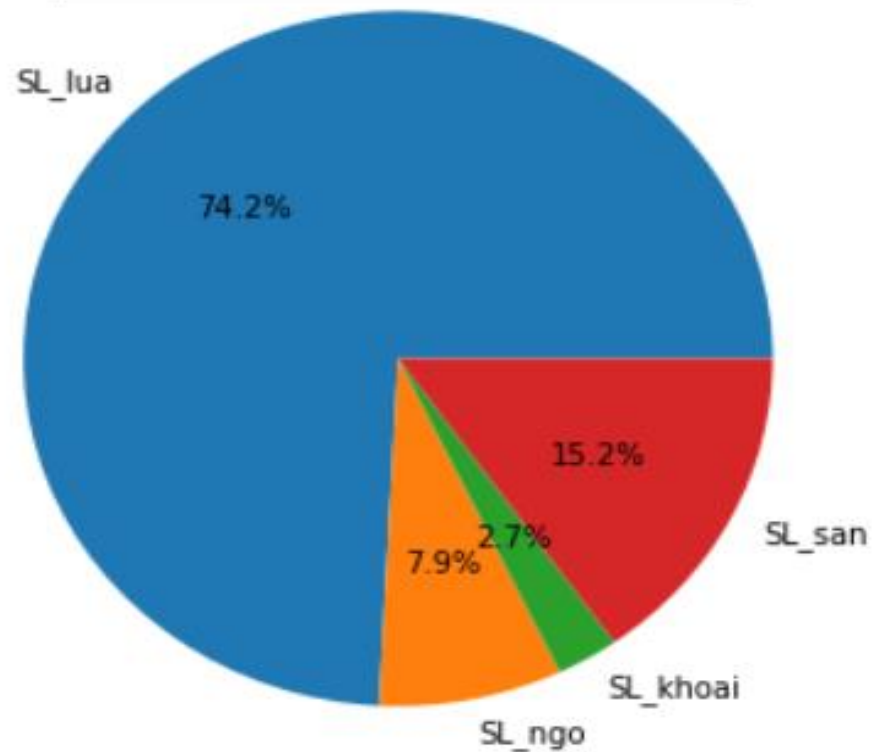


## So sánh tỉ trọng diện tích và sản lượng cây lương thực

Tỉ trọng diện tích cây lương thực

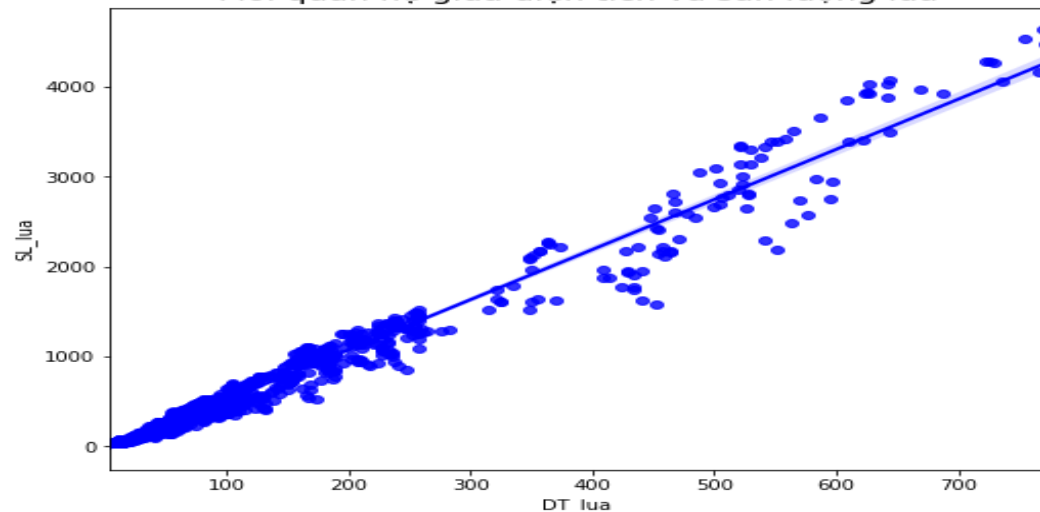


Tỉ trọng sản lượng cây lương thực

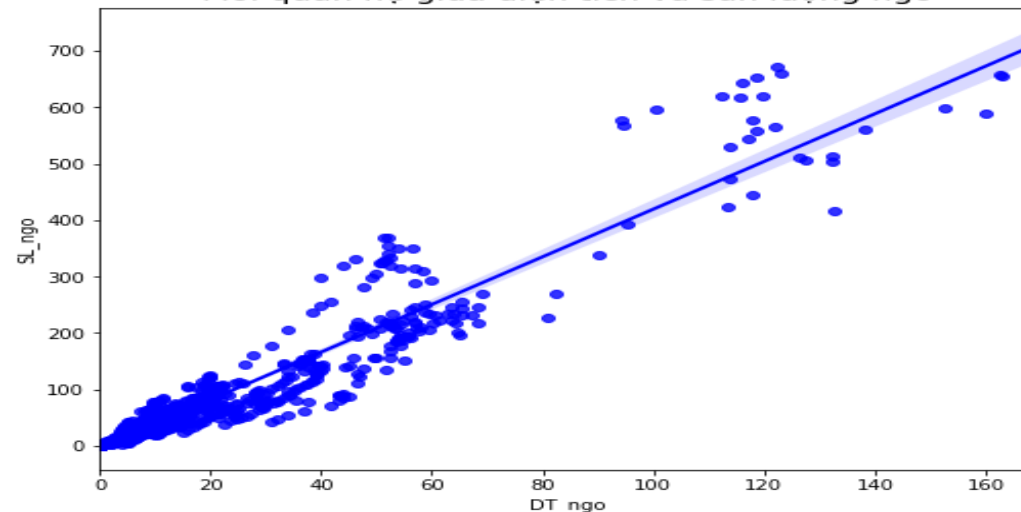


# Mối quan hệ giữa diện tích và sản lượng của các cây lương thực

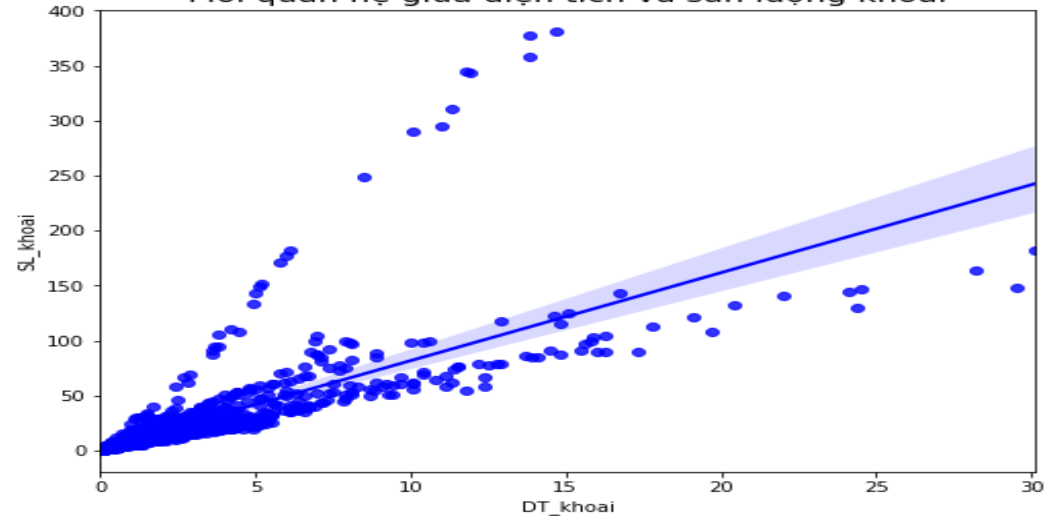
Mối quan hệ giữa diện tích và sản lượng lúa



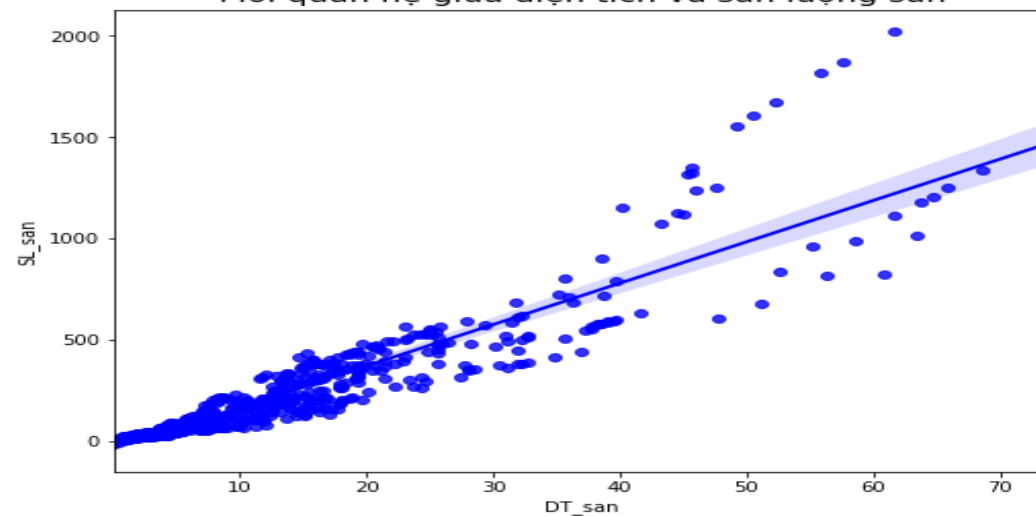
Mối quan hệ giữa diện tích và sản lượng ngô



Mối quan hệ giữa diện tích và sản lượng khoai



Mối quan hệ giữa diện tích và sản lượng sắn



# KIỂM ĐỊNH SỐ LIỆU GIỮA DIỆN TÍCH VÀ SẢN LƯỢNG

## 1) Lúa

### OLS Regression Results

```
=====
Dep. Variable:          SL_lua      R-squared:                0.972
Model:                  OLS         Adj. R-squared:            0.972
Method:                 Least Squares   F-statistic:           4.332e+04
Date:                  Sun, 18 Jul 2021   Prob (F-statistic):      0.00
Time:                  02:16:20         Log-Likelihood:        -7822.4
No. Observations:      1248           AIC:                  1.565e+04
Df Residuals:          1246           BIC:                  1.566e+04
Df Model:               1
Covariance Type:       nonrobust
=====
```

```
=====
              coef      std err          t      P>|t|      [0.025      0.975]
-----
const      -44.2187      4.841      -9.133      0.000     -53.717     -34.720
DT_lua       5.5809      0.027     208.141      0.000       5.528       5.633
=====
```

```
=====
Omnibus:            411.667   Durbin-Watson:           0.922
Prob(Omnibus):       0.000   Jarque-Bera (JB):       5044.936
Skew:                -1.167   Prob(JB):               0.00
Kurtosis:            12.569   Cond. No.               242.
=====
```

### • Nhận xét:

- Biến DT\_lua có ý nghĩa về mặt thống kê.
- Mô hình phù hợp tốt với dữ liệu quan sát về mặt thống kê.
- Biến DT\_lua có thể giải thích được 97.2% sự thay đổi của biến SL\_lua.

### - Phương trình hồi quy:

$$SL\_lua = -44.2187 + 5.5809 * DT\_lua$$



# KIỂM ĐỊNH SỐ LIỆU GIỮA DIỆN TÍCH VÀ SẢN LƯỢNG

## 2) Ngô

### OLS Regression Results

Dep. Variable:	SL_ngo	R-squared:	0.917			
Model:	OLS	Adj. R-squared:	0.917			
Method:	Least Squares	F-statistic:	1.367e+04			
Date:	Sun, 18 Jul 2021	Prob (F-statistic):	0.00			
Time:	02:16:20	Log-Likelihood:	-5978.4			
No. Observations:	1243	AIC:	1.196e+04			
Df Residuals:	1241	BIC:	1.197e+04			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	-2.9094	1.035	-2.811	0.005	-4.940	-0.879
DT_ngo	4.2245	0.036	116.917	0.000	4.154	4.295
=====						
Omnibus:	505.491	Durbin-Watson:	1.147			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6207.285			
Skew:	1.532	Prob(JB):	0.00			
Kurtosis:	13.510	Cond. No.	35.2			
=====						

### • Nhận xét:

- Biến DT\_ngo có ý nghĩa về mặt thống kê.
- Mô hình phù hợp tốt với dữ liệu quan sát về mặt thống kê.
- Biến DT\_ngo có thể giải thích được 91.7% sự thay đổi của biến SL\_ngo.

### - Phương trình hồi quy:

$$SL\_ngo = -2.9094 + 4.2245 * DT\_ngo$$

# KIỂM ĐỊNH SỐ LIỆU GIỮA DIỆN TÍCH VÀ SẢN LƯỢNG

## 3) Khoai

### OLS Regression Results

```
=====
Dep. Variable:          SL_khoai    R-squared:                0.575
Model:                  OLS         Adj. R-squared:            0.575
Method:                 Least Squares   F-statistic:             1673.
Date:                  Sun, 18 Jul 2021   Prob (F-statistic):      5.10e-232
Time:                  02:54:50         Log-Likelihood:          -5719.2
No. Observations:      1238           AIC:                    1.144e+04
Df Residuals:          1236           BIC:                    1.145e+04
Df Model:               1
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	1.5985	0.866	1.845	0.065	-0.101	3.298
DT_khoai	8.0003	0.196	40.908	0.000	7.617	8.384

```
=====
Omnibus:                1584.312    Durbin-Watson:           1.410
Prob(Omnibus):           0.000      Jarque-Bera (JB):        209651.858
Skew:                    6.863      Prob(JB):                0.00
Kurtosis:                65.257     Cond. No.                 5.60
=====
```

### • Nhận xét:

- Biến DT\_khoai có ý nghĩa về mặt thống kê.
- Mô hình phù hợp tốt với dữ liệu quan sát về mặt thống kê.
- Biến DT\_khoai có thể giải thích được 57.5% sự thay đổi của biến SL\_khoai.

- Phương trình hồi quy:

$$SL\_khoai = 1.5985 + 8.0003 * DT\_khoai$$

# KIỂM ĐỊNH SỐ LIỆU GIỮA DIỆN TÍCH VÀ SẢN LƯỢNG

## 4) Sản

### OLS Regression Results

```
=====
Dep. Variable:          SL_san      R-squared:          0.864
Model:                  OLS         Adj. R-squared:       0.864
Method:                 Least Squares   F-statistic:        4999.
Date:                  Sun, 18 Jul 2021   Prob (F-statistic):  0.00
Time:                  02:57:30         Log-Likelihood:     -4733.4
No. Observations:      788            AIC:               9471.
Df Residuals:          786            BIC:               9480.
Df Model:              1
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	-37.9813	4.863	-7.810	0.000	-47.527	-28.435
DT_san	20.4094	0.289	70.704	0.000	19.843	20.976

```
=====
Omnibus:              524.671      Durbin-Watson:       1.363
Prob(Omnibus):        0.000        Jarque-Bera (JB):    13930.046
Skew:                 2.580        Prob(JB):           0.00
Kurtosis:             22.941       Cond. No.           23.4
=====
```

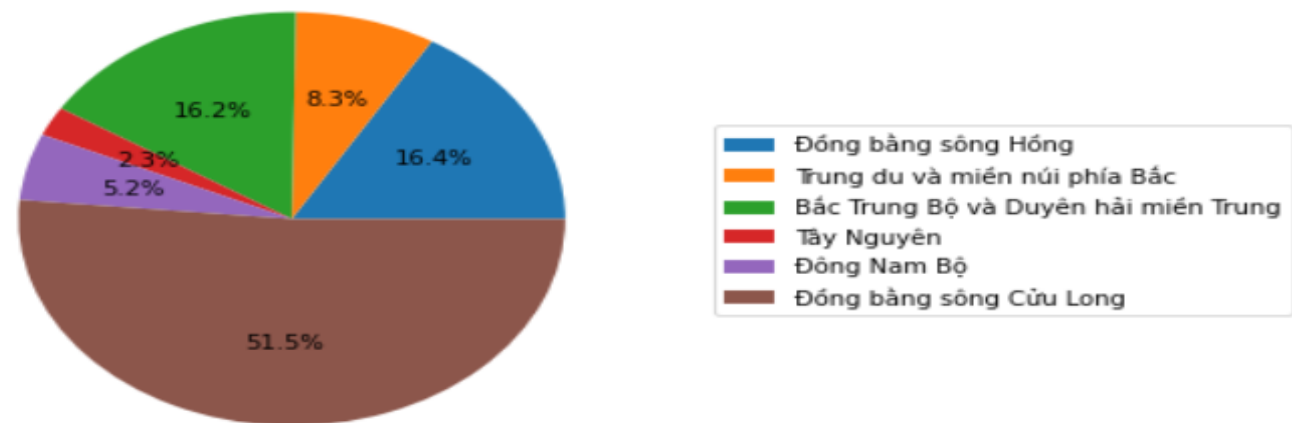
### • Nhận xét:

- Biến DT\_sản có ý nghĩa về mặt thống kê.
- Mô hình phù hợp tốt với dữ liệu quan sát về mặt thống kê.
- Biến DT\_sản có thể giải thích được 86.4% sự thay đổi của biến SL\_sản.
- Phương trình hồi quy:

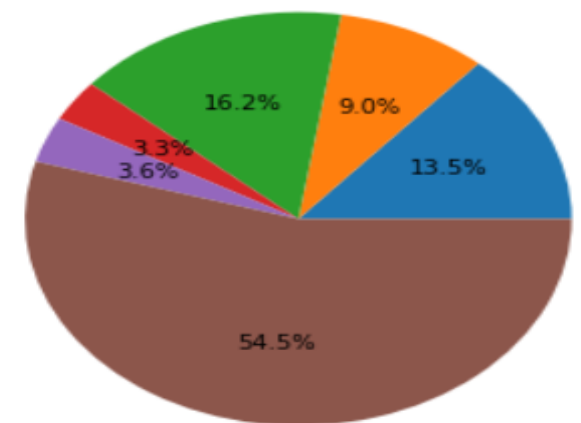
$$SL_{\text{sản}} = -37.9813 + 20.4094 * DT_{\text{sản}}$$

# So sánh tỉ trọng trồng cây lương thực ở các khu vực trong năm 2000 và năm 2019

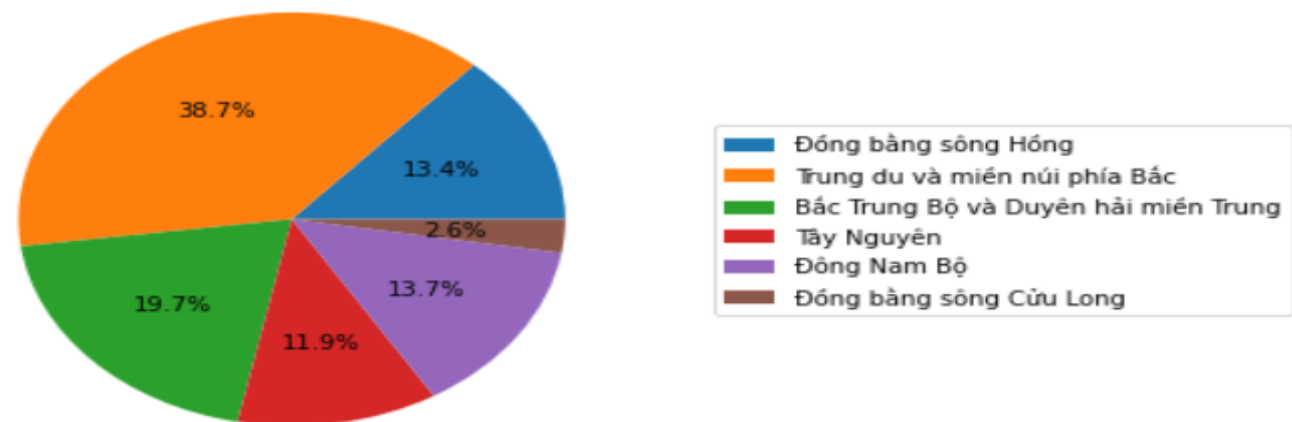
Diện tích lúa theo khu vực năm 2000



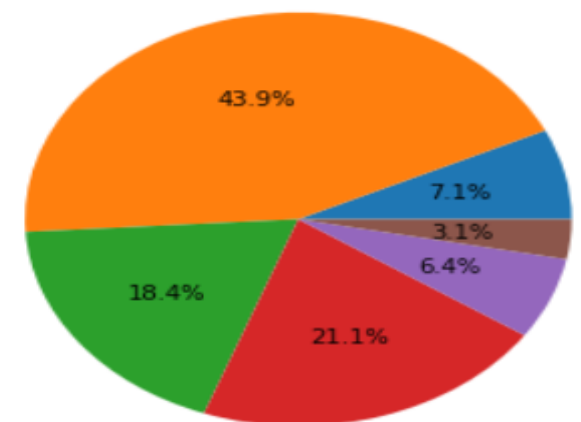
Diện tích lúa theo khu vực năm 2019



Diện tích ngô theo khu vực năm 2000

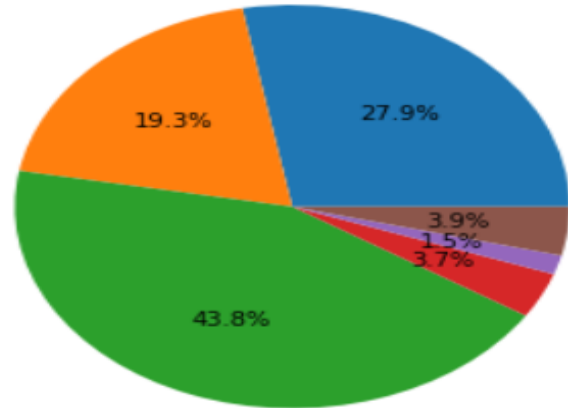


Diện tích ngô theo khu vực năm 2019



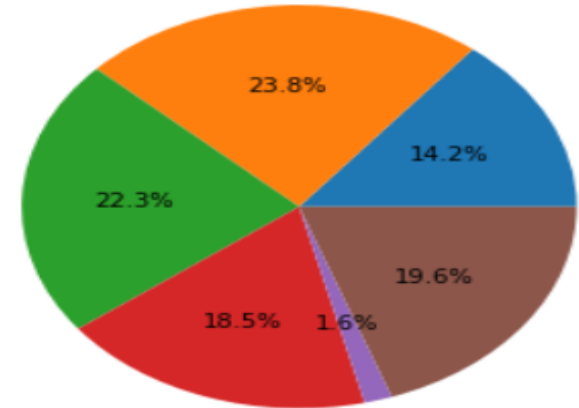
# So sánh tỉ trọng trồng cây lương thực ở các khu vực trong năm 2000 và năm 2019

Diện tích khoai theo khu vực năm 2000

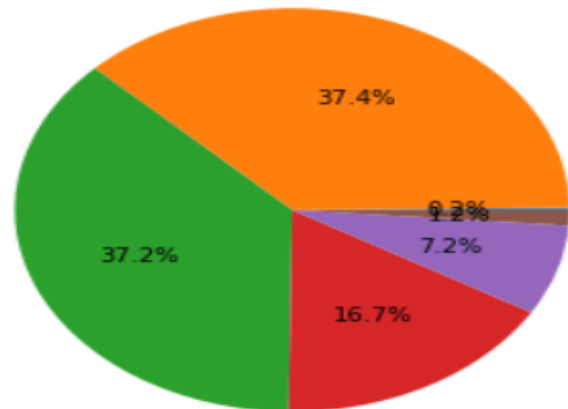


- Đồng bằng sông Hồng
- Trung du và miền núi phía Bắc
- Bắc Trung Bộ và Duyên hải miền Trung
- Tây Nguyên
- Đông Nam Bộ
- Đồng bằng sông Cửu Long

Diện tích khoai theo khu vực năm 2019

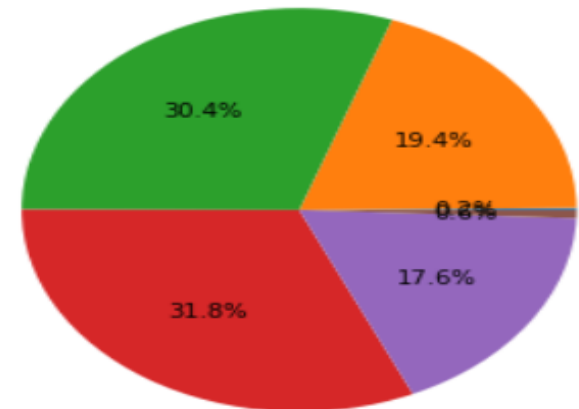


Diện tích sắn theo khu vực năm 2000

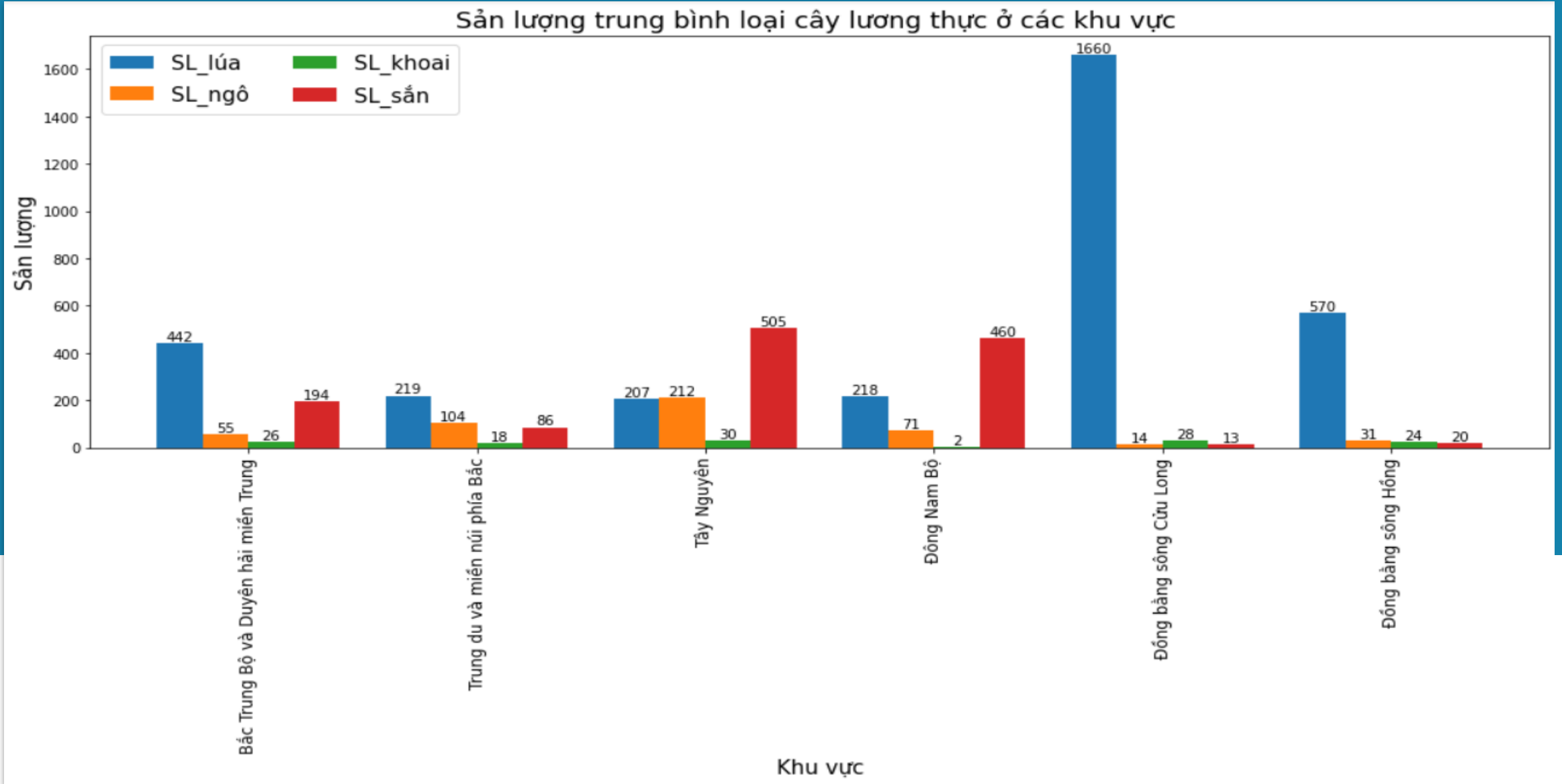


- Đồng bằng sông Hồng
- Trung du và miền núi phía Bắc
- Bắc Trung Bộ và Duyên hải miền Trung
- Tây Nguyên
- Đông Nam Bộ
- Đồng bằng sông Cửu Long

Diện tích sắn theo khu vực năm 2019



# Sản lượng trung bình các loại cây lương thực ở các khu vực

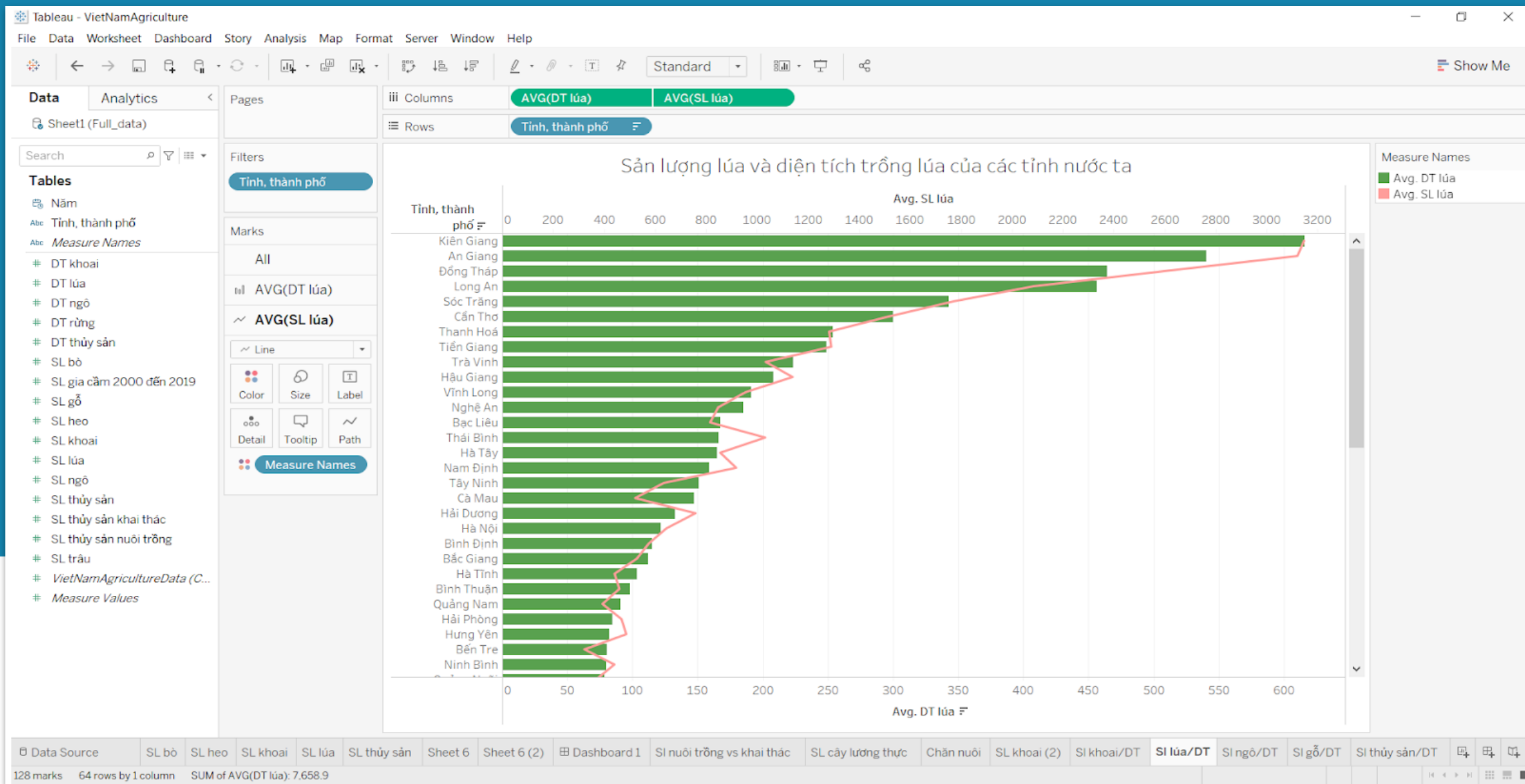


# ƯỚC LƯỢNG KHOẢNG TIN CÂY

Với mức ý nghĩa 0.05, ta tìm khoảng tin cậy cho sản lượng và diện tích của tất cả các loại cây lương thực.

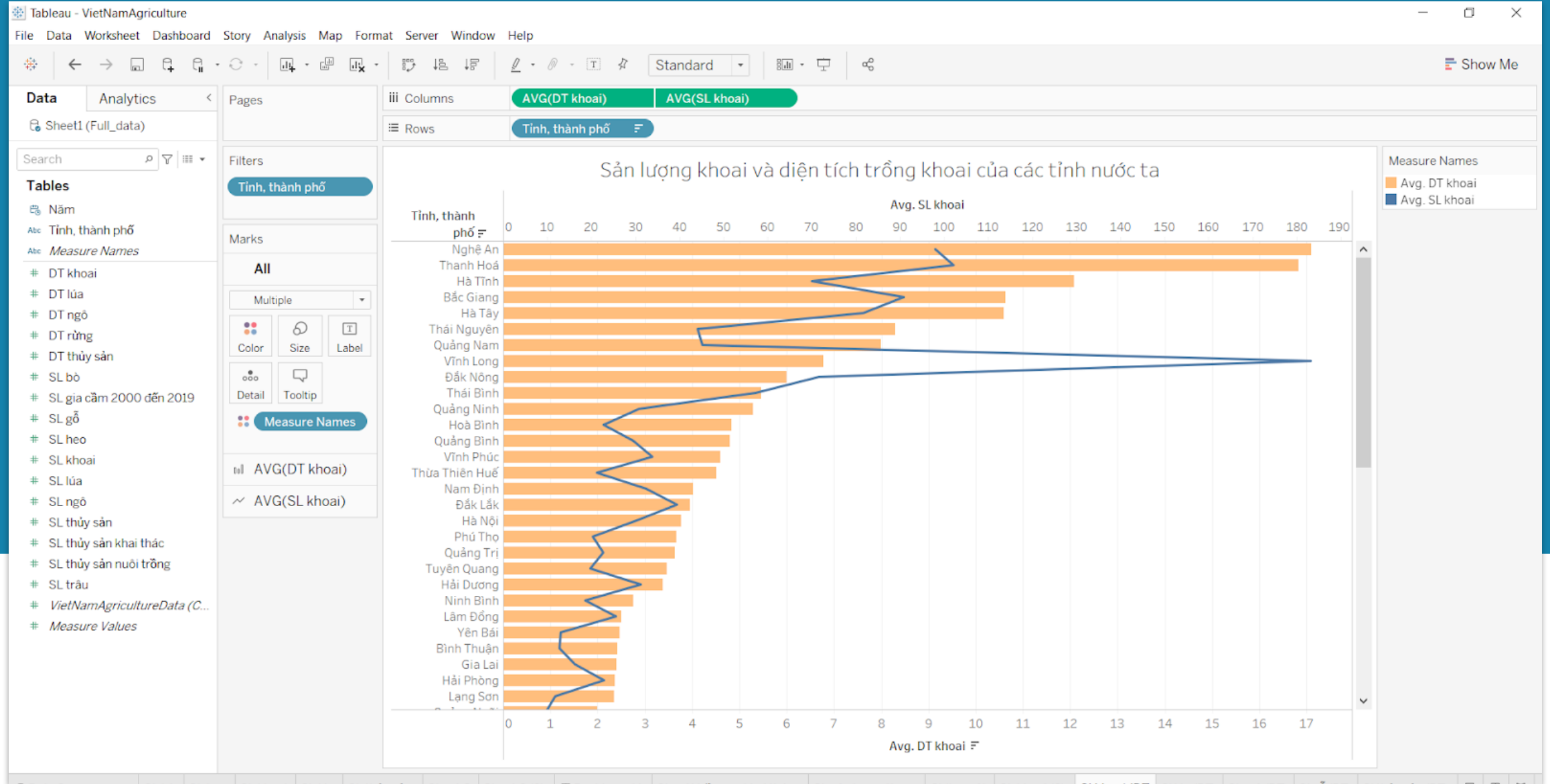
	Giới hạn dưới	Giới hạn trên
DT_lua	1029.383985	1490.661015
SL_lua	5307.356946	7838.521388
DT_ngo	149.169701	197.471965
SL_ngo	614.391489	789.583511
DT_khoai	23.449110	31.844223
SL_khoai	209.244496	266.398837
DT_san	65.937100	87.401233
SL_san	1114.617429	1519.852571

# Sản lượng và diện tích lúa theo tỉnh





# Sản lượng và diện tích khoai theo tỉnh



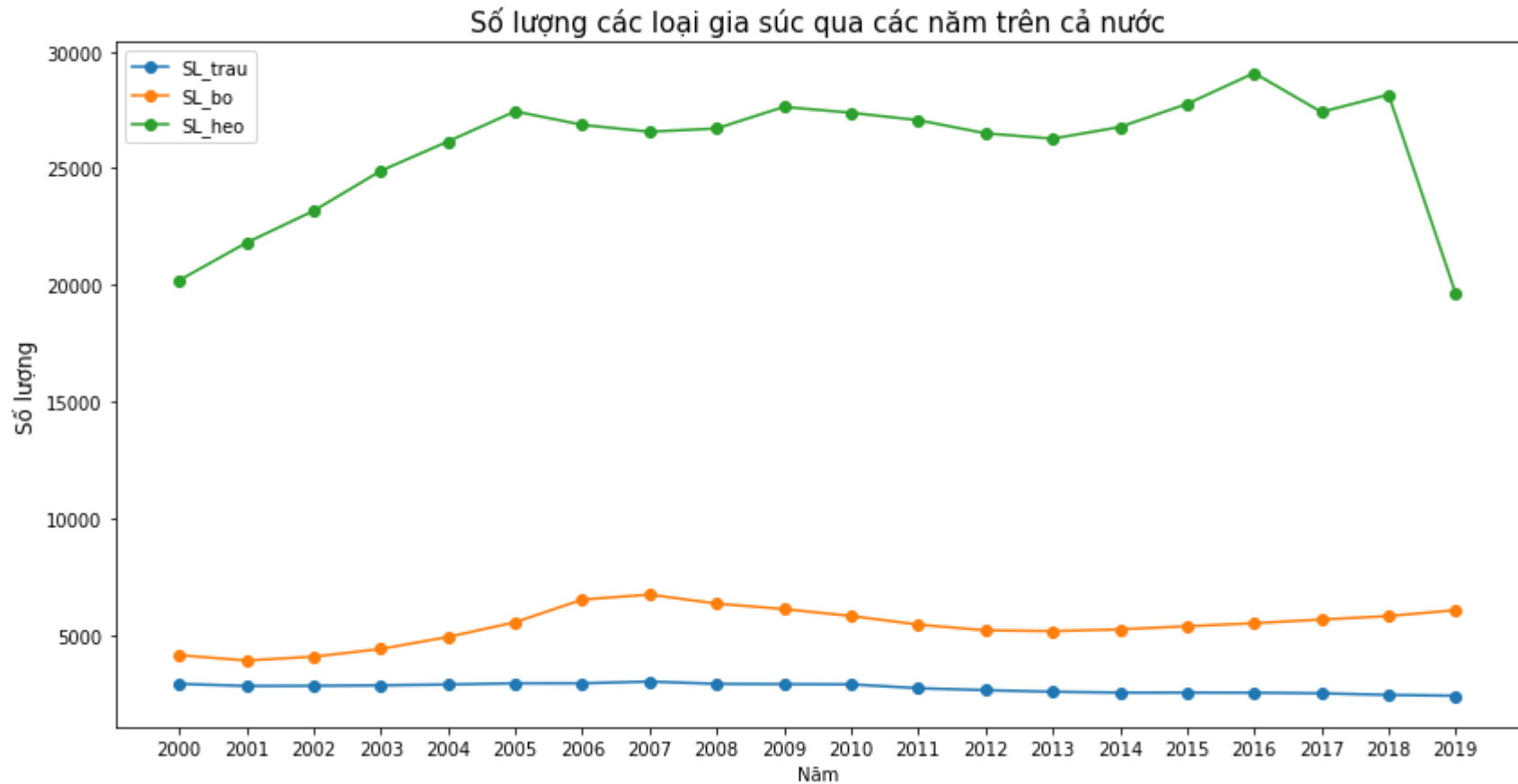
## 4. TRỰC QUAN VÀ PHÂN TÍCH DỮ LIỆU CHĂN NUÔI

Dữ liệu chăn nuôi bao gồm:

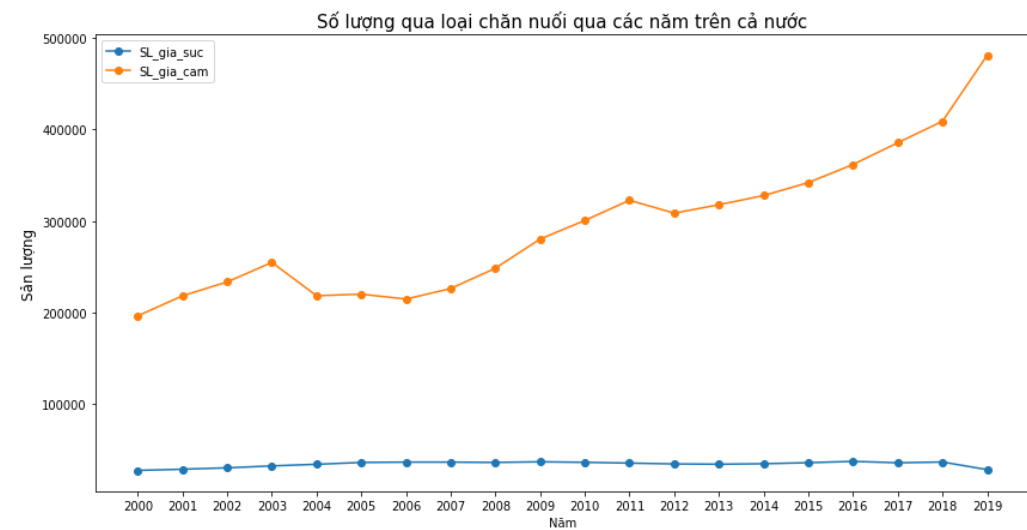
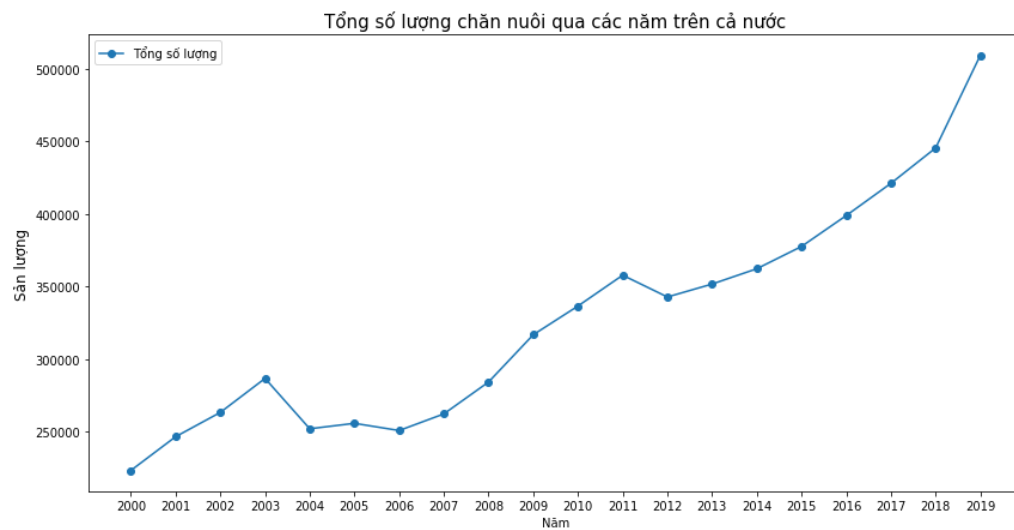
- Số lượng (nghìn con)

của gia súc (trâu, bò, heo) và gia cầm trên một đơn vị diện tích lãnh thổ (tỉnh, khu vực, cả nước) trong 1 năm trong khoảng 2000-2019

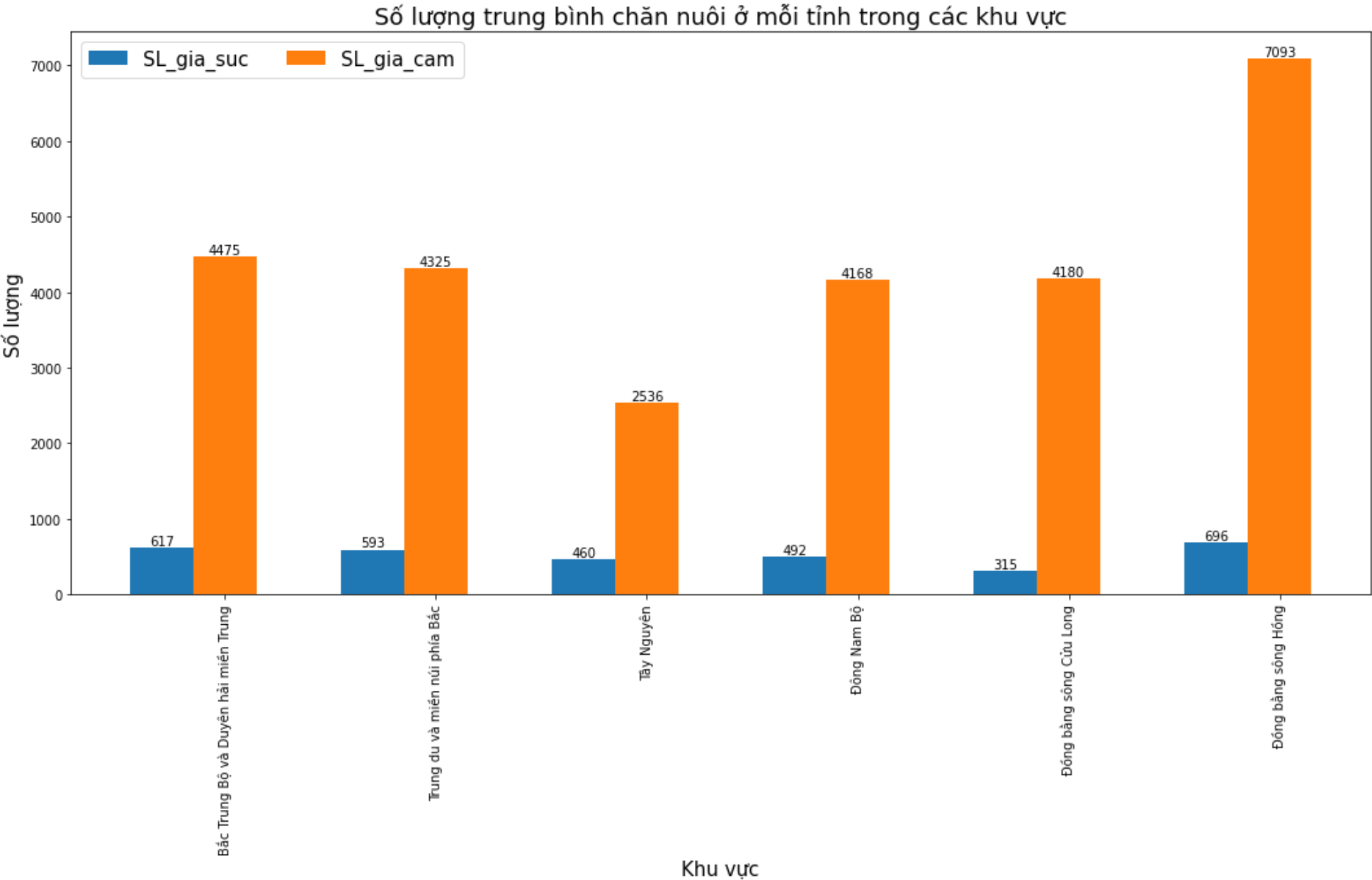
# So sánh số lượng trâu, bò, heo của cả nước qua các năm



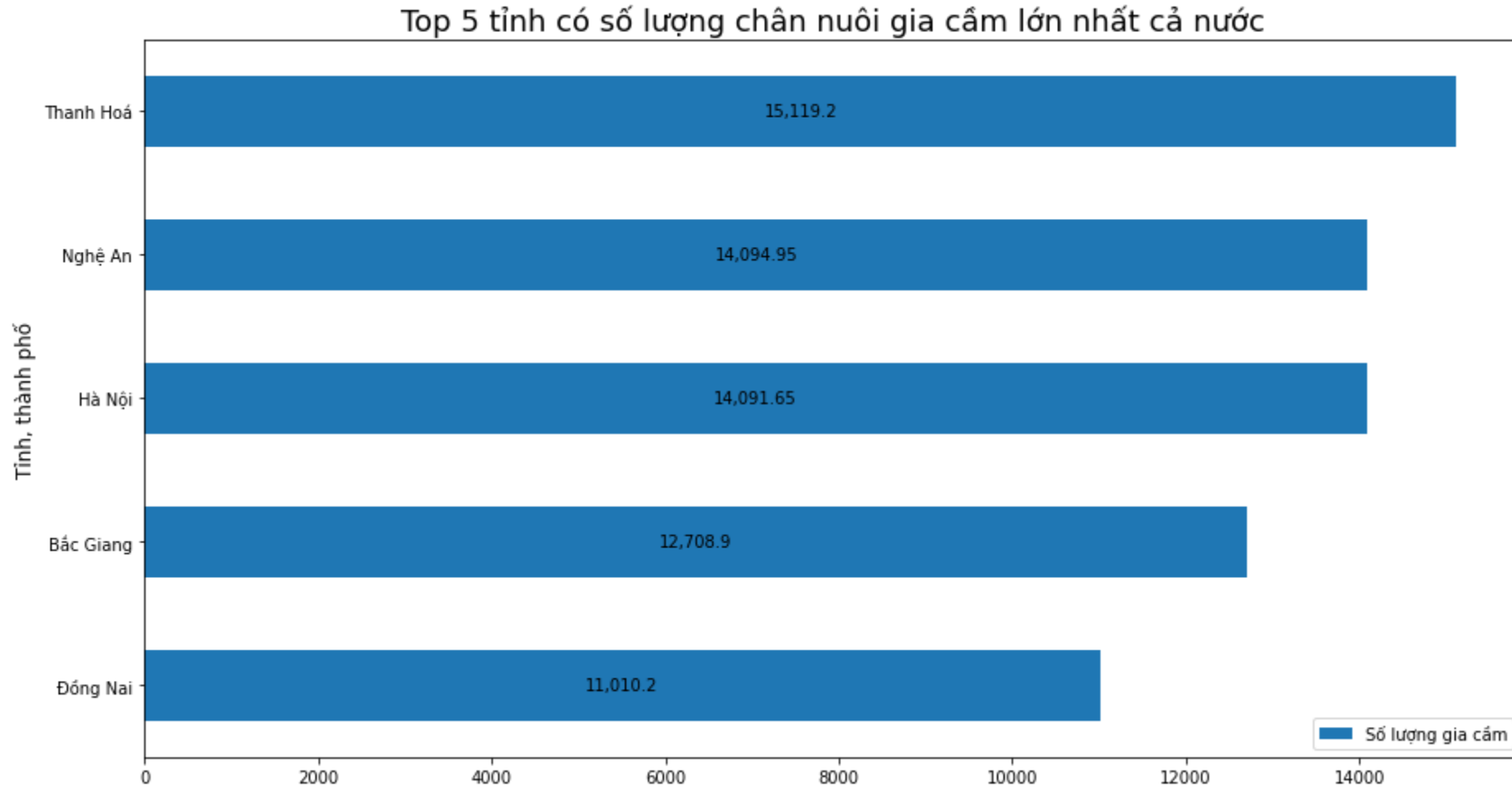
# So sánh biến động tổng số lượng chăn nuôi của cả nước qua các năm



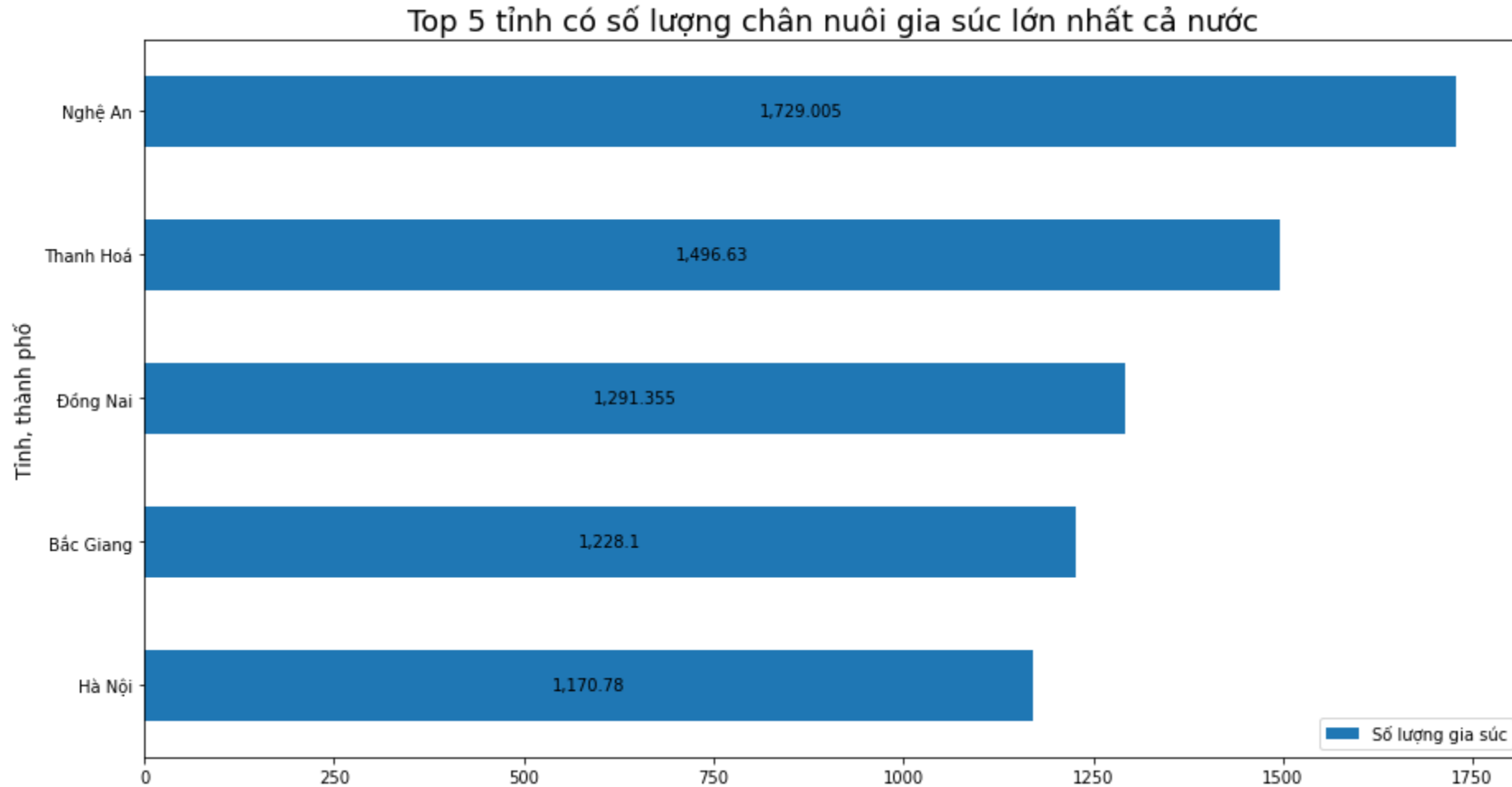
# Số lượng chăn nuôi trung bình của các khu vực



# Top 5 tỉnh đứng đầu cả nước về số lượng chăn nuôi gia cầm



# Top 5 tỉnh đứng đầu cả nước về số lượng chăn nuôi gia súc



# MÔ HÌNH HỒI QUY BIỂU DIỄN MỐI QUAN HỆ GIỮA SẢN LƯỢNG GIA SÚC VÀ GIA CẦM

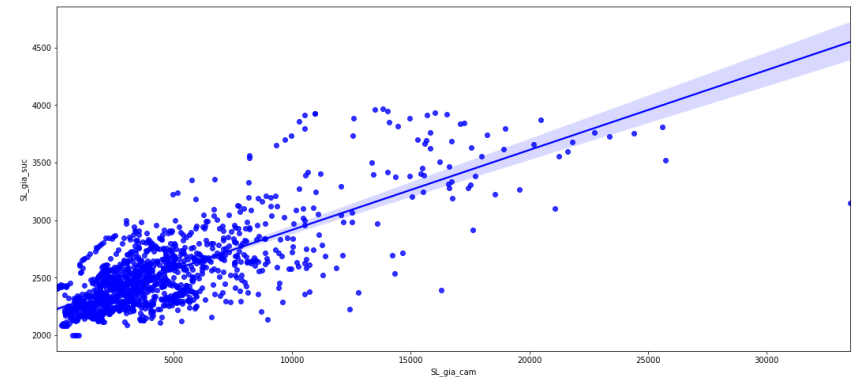
## Kết luận:

- Biến SL\_gia\_cam có ý nghĩa đối với mô hình về mặt thống kê (p-value = 0).
- Mô hình phù hợp tốt với dữ liệu quan sát về mặt thống kê.
- Biến SL\_gia\_suc có thể giải thích được 58.5% sự thay đổi của biến SL\_gia\_cam.
- Phương trình hồi quy:

$$SL_{giasuc} = 2223.1530 + 0.0693 \times SL_{gia\_cam}$$

Khi số lượng gia súc tăng 10 con thì số lượng gia cầm tăng 145 con.

OLS Regression Results						
=====						
Dep. Variable:	SL_gia_suc	R-squared:	0.585			
Model:	OLS	Adj. R-squared:	0.585			
Method:	Least Squares	F-statistic:	1755.			
Date:	Sun, 18 Jul 2021	Prob (F-statistic):	4.08e-240			
Time:	06:50:33	Log-Likelihood:	-8551.1			
No. Observations:	1248	AIC:	1.711e+04			
Df Residuals:	1246	BIC:	1.712e+04			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	2223.1530	10.044	221.352	0.000	2203.449	2242.857
SL_gia_cam	0.0693	0.002	41.897	0.000	0.066	0.073
=====						
Omnibus:	120.833	Durbin-Watson:	1.135			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	466.661			
Skew:	0.399	Prob(JB):	4.63e-102			
Kurtosis:	5.888	Cond. No.	9.40e+03			
=====						

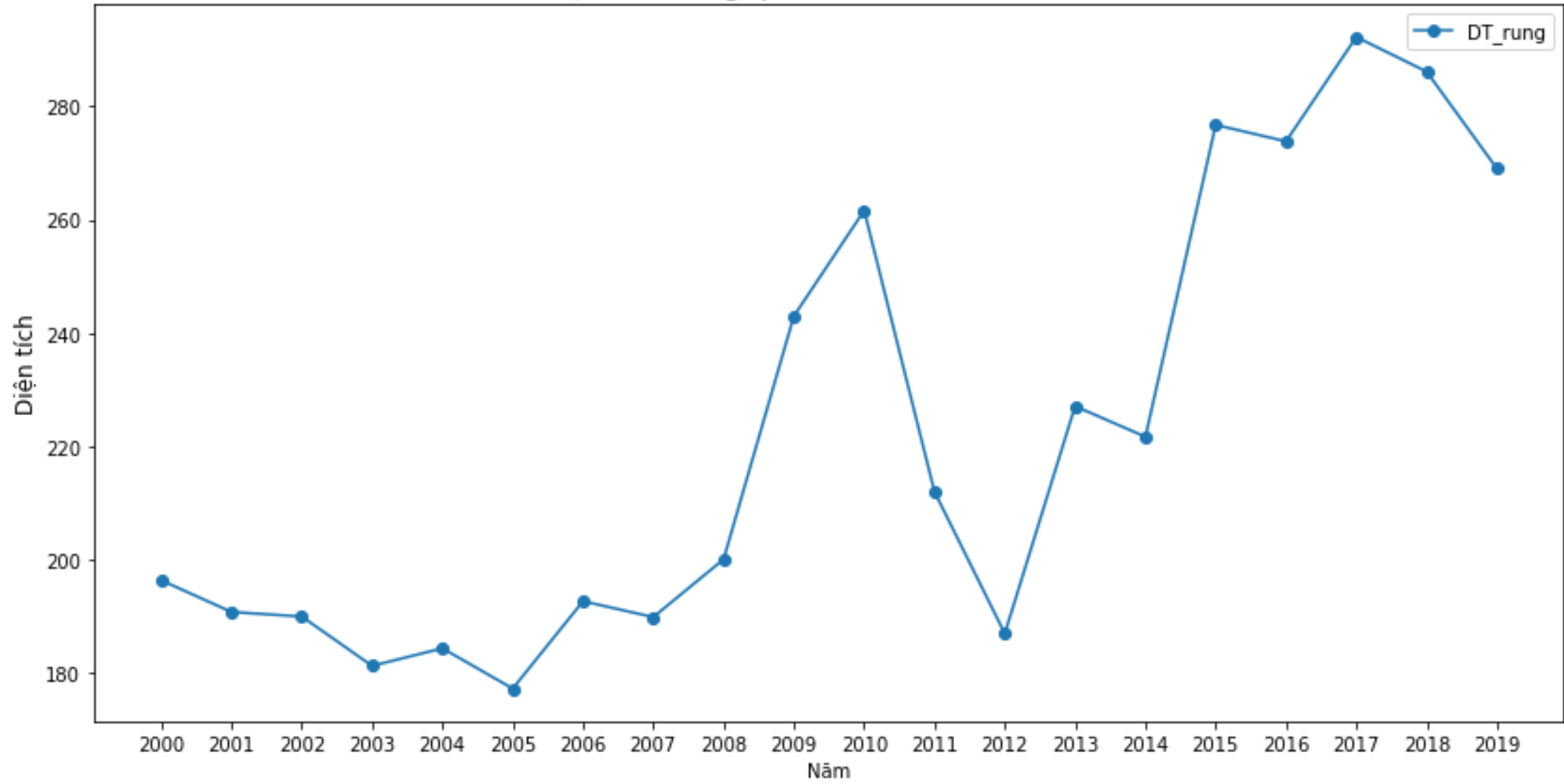




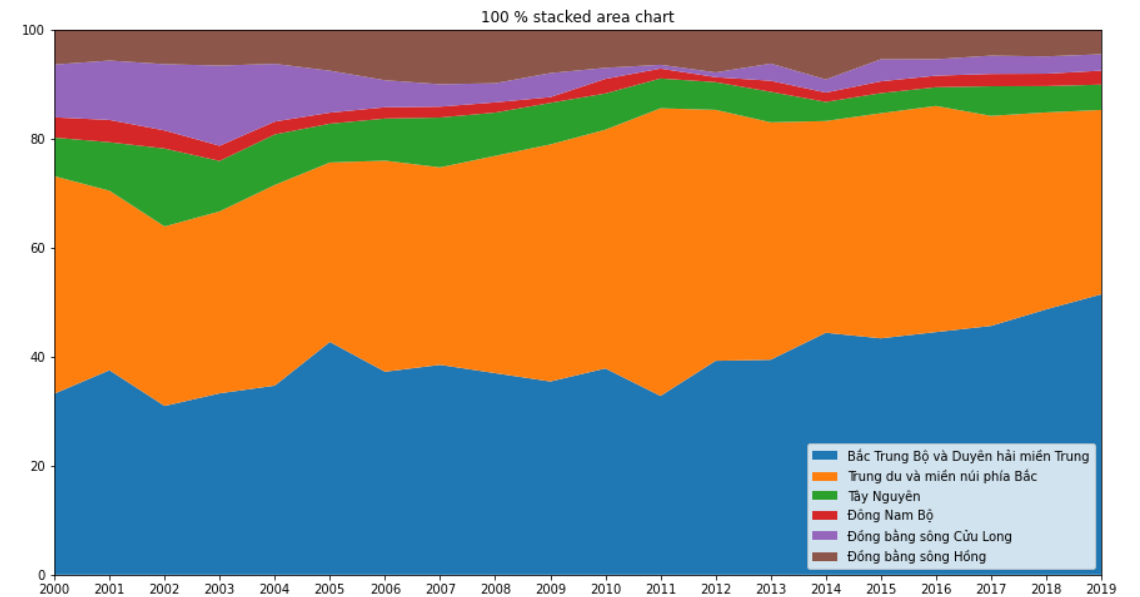
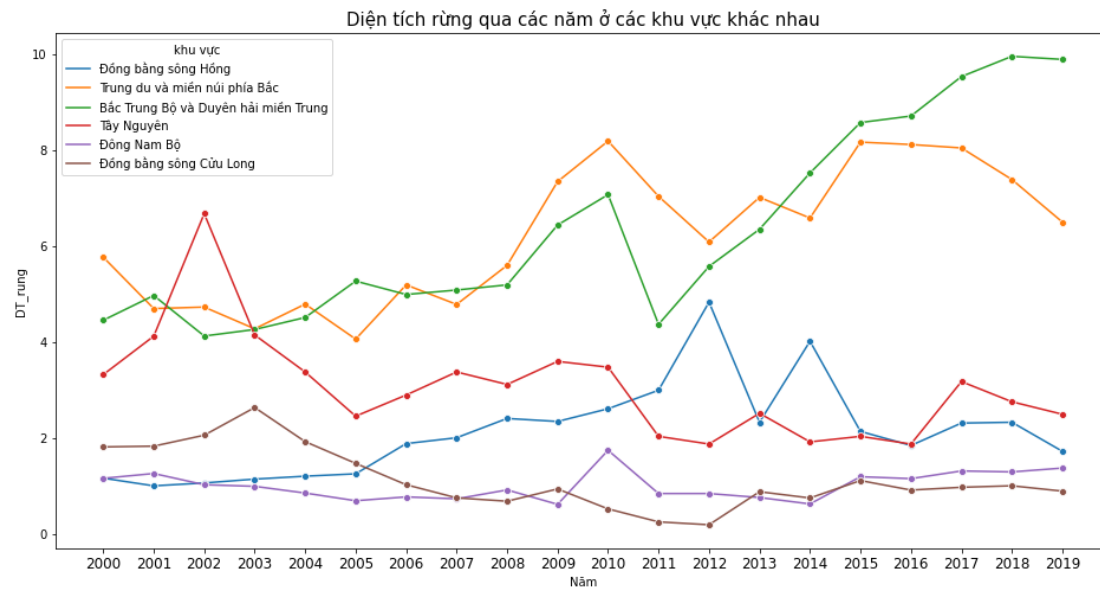
## 5. TRỰC QUAN VÀ PHÂN TÍCH DỮ LIỆU RỪNG

Dữ liệu rừng bao gồm diện tích rừng (nghìn ha) và sản lượng (nghìn tấn) trên một đơn vị diện tích lãnh thổ (tỉnh, khu vực, cả nước) trong 1 năm trong khoảng 2000-2019

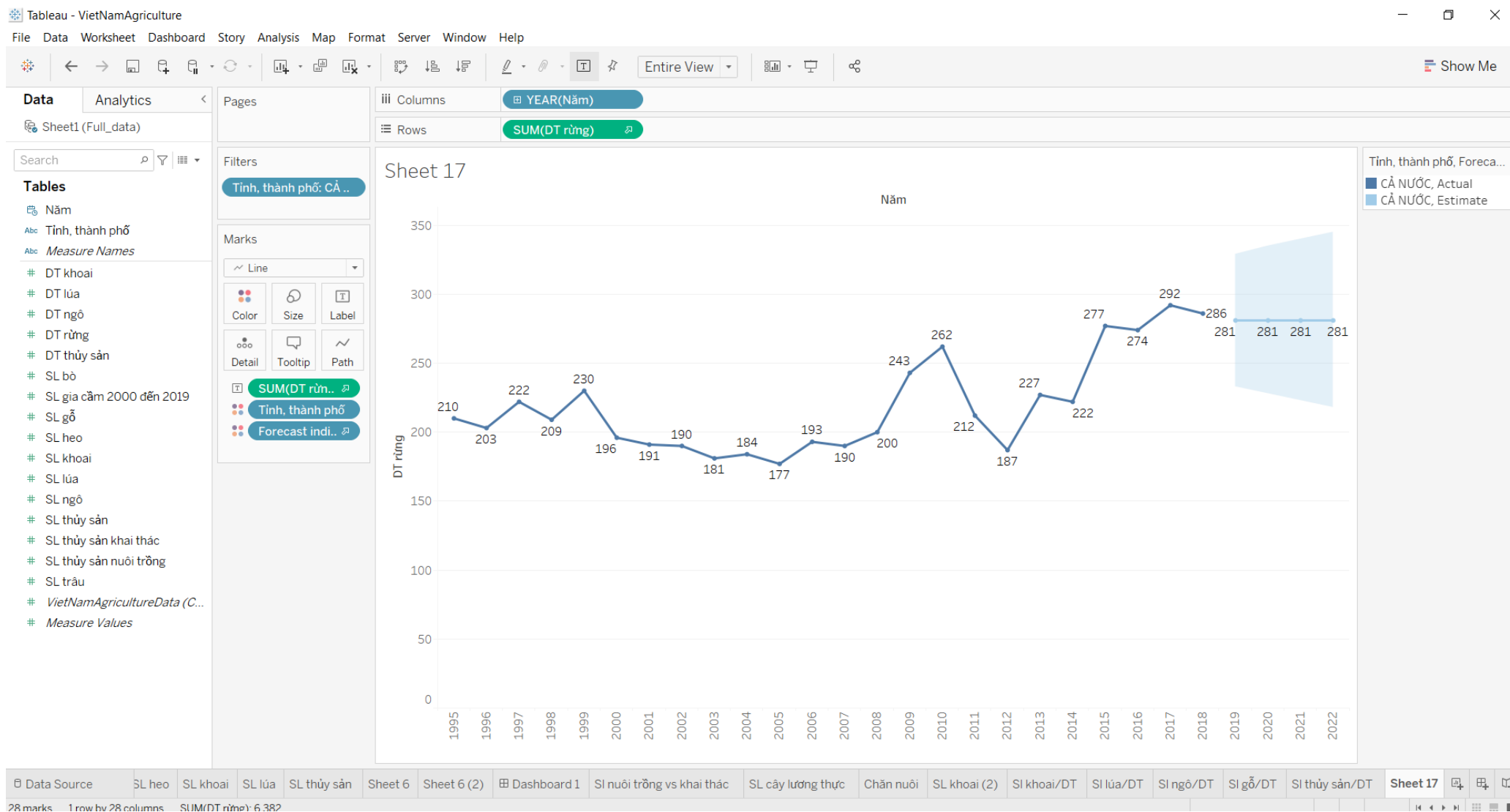
Diện tích rừng qua các năm trên cả nước



# Diện tích rừng và tỷ trọng của các khu vực qua các năm



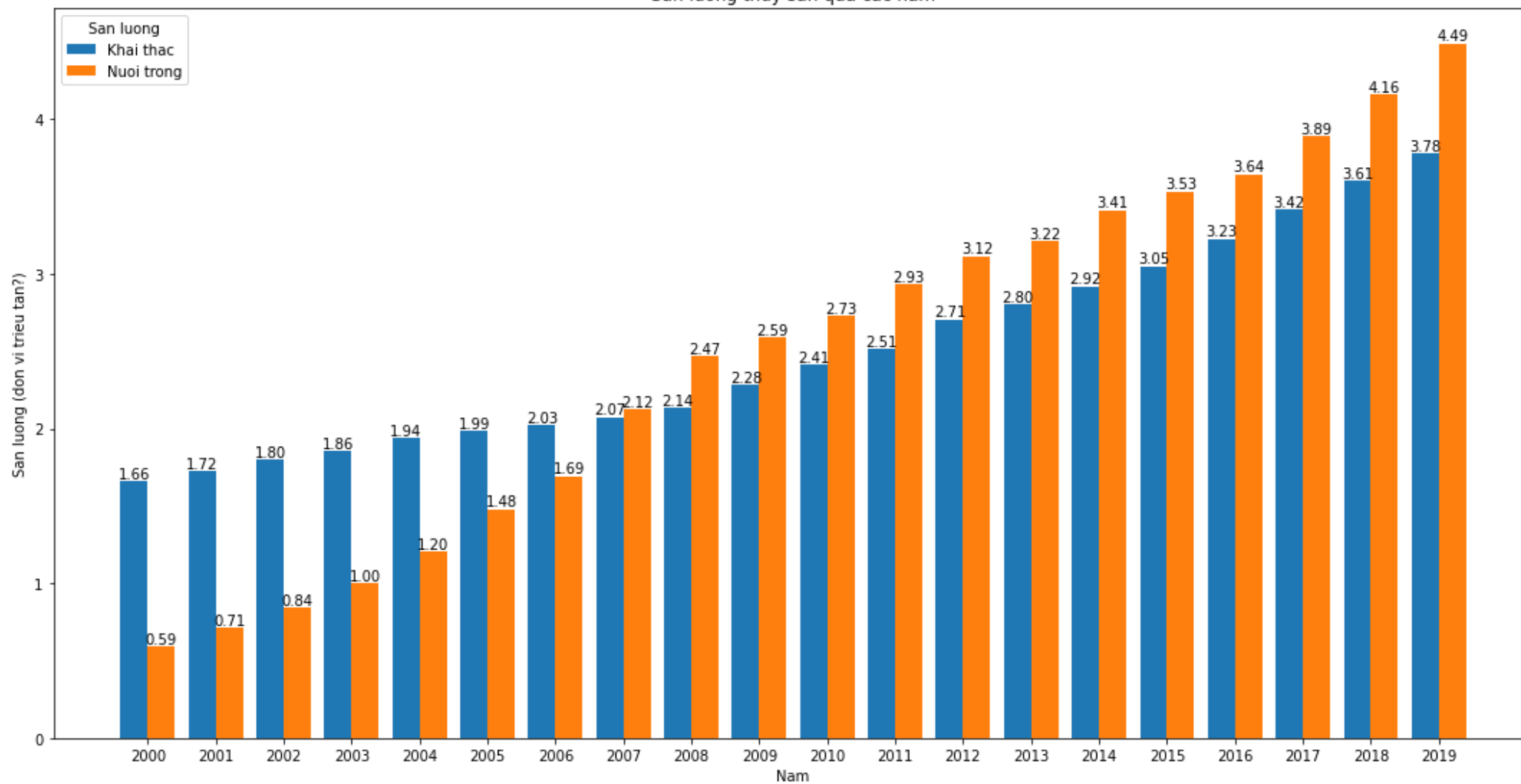
# Sử dụng tableau để dự đoán sản lượng gỗ của cả nước trong tương lai gần



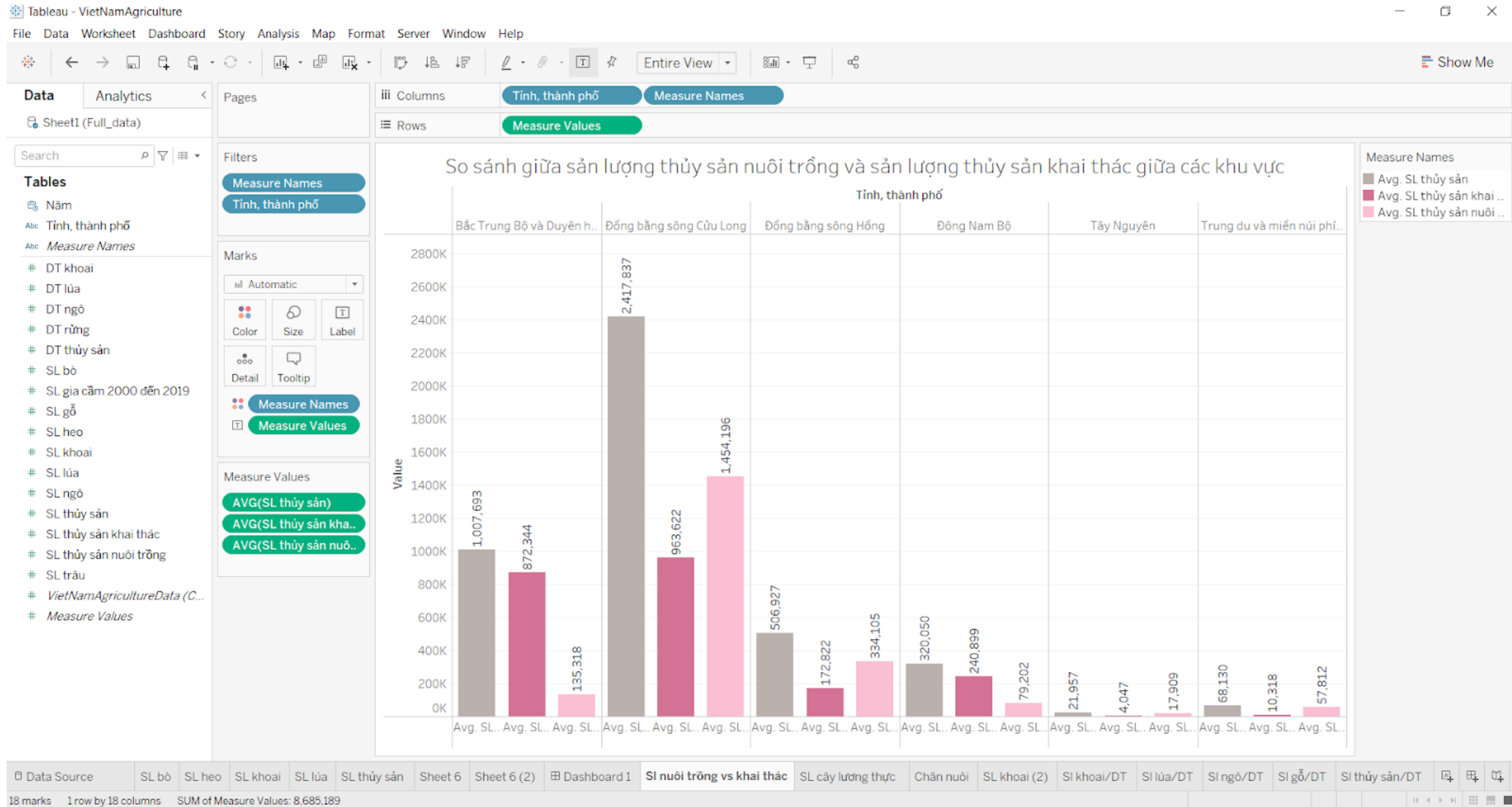
## 6. TRỰC QUAN VÀ PHÂN TÍCH DỮ LIỆU NGƯ NGHIỆP

Dữ liệu ngư nghiệp bao gồm sản lượng thủy sản khai thác và nuôi trồng (nghìn tấn) trên một đơn vị diện tích lãnh thổ (tỉnh, khu vực, cả nước) trong 1 năm trong khoảng 2000-2019

San lượng thủy sản qua các năm



# So sánh sản lượng thủy sản nuôi trồng và khai thác giữa các khu vực



# MÔ HÌNH HỒI QUY BIỂU DIỄN MỐI QUAN HỆ GIỮA SẢN LƯỢNG THỦY SẢN NUÔI TRỒNG VÀ KHAI THÁC QUA CÁC NĂM

## Kết luận:

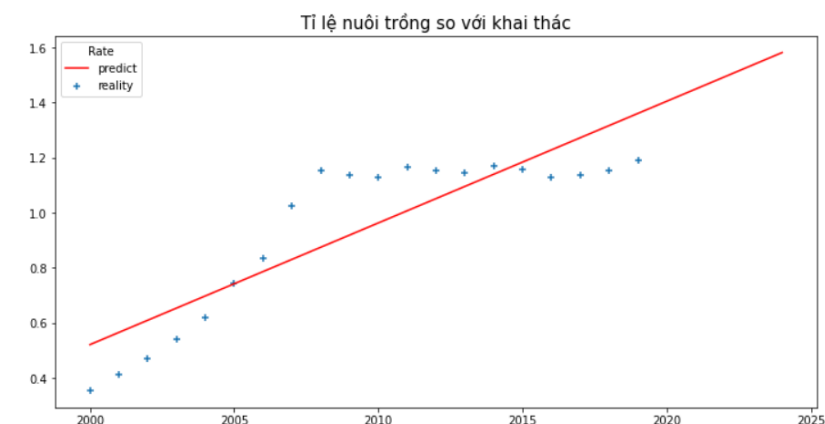
- Biến X có ý nghĩa đối với mô hình về mặt thống kê (p-value = 0).
- Mô hình phù hợp tốt với dữ liệu quan sát về mặt thống kê.
- Biến y có thể giải thích được 76.2% sự thay đổi của biến X.
- Phương trình hồi quy:

Tỉ lệ sản lượng thủy sản nuôi trồng so với khai thác =  $0.5214 + 0.0442 \times \text{Năm}$

## Theo kết quả của mô hình:

- Tỉ lệ tăng 0.045 khi cứ qua mỗi năm.

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	0.762			
Model:	OLS	Adj. R-squared:	0.748			
Method:	Least Squares	F-statistic:	57.53			
Date:	Sun, 18 Jul 2021	Prob (F-statistic):	5.19e-07			
Time:	06:50:35	Log-Likelihood:	10.590			
No. Observations:	20	AIC:	-17.18			
Df Residuals:	18	BIC:	-15.19			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	0.5214	0.065	8.055	0.000	0.385	0.657
x1	0.0442	0.006	7.585	0.000	0.032	0.056
=====						
Omnibus:	3.283	Durbin-Watson:	0.154			
Prob(Omnibus):	0.194	Jarque-Bera (JB):	1.673			
Skew:	0.404	Prob(JB):	0.433			
Kurtosis:	1.836	Cond. No.	21.5			





# Dự đoán sản lượng thủy sản trong tương lai gần của 10 tỉnh có sản lượng thủy sản đứng đầu cả nước

