



# Working Hours Analysis

---

Zesen Huang

Yinxia Shen

Jiani Wu



# CONTENTS

**01**

**Data describe**

**02**

**Overview of data**

**03**

**Data analysis and Model**

**04**

**Conclusion**

# **/01 Data describe**



# Data describe

---

- This subject is about studying what factors related to working hours in different countries.
- The data comes from Our World in Data(<https://ourworldindata.org/working-hours>). Two copies of the initial data were downloaded from the site, which are the annual work duration against GDP, annual working duration against productivity.

Data of GDP	Entity	Code	Year	Average annual hours worked	GDP	Population	Continent
Data of productivity	Entity	Code	Year	Average annual hours worked	Productivity	Population	Continent

# **/02 Overview of data**



# Overview of data

---

- After cleaning and filtering data, we get the following data:

	Entity	Year	Average annual working hours	GDP	Population	Continent	Productivity
1	Argentina	2010	1751.765	18712.063	40896000	South America	20.230000
2	Argentina	2011	1750.974	19629.352	41320000	South America	21.449110
3	Argentina	2012	1726.258	19224.874	41755000	South America	21.437298
4	Argentina	2013	1714.871	19482.190	42196000	South America	21.548719
5	Argentina	2014	1695.364	18797.548	42638000	South America	21.150030
6	Argentina	2015	1691.536	19116.908	43075000	South America	21.299370
7	Argentina	2016	1691.536	18584.580	43508000	South America	21.183163
8	Argentina	2017	1691.536	18933.907	43937000	South America	21.749641
9	Australia	2010	1769.309	41464.040	22155000	Oceania	50.503040
10	Australia	2011	1768.649	41894.173	22538000	Oceania	52.322773
11	Australia	2012	1759.519	42754.456	22904000	Oceania	52.503056
12	Australia	2013	1755.827	43118.086	23255000	Oceania	53.549725
13	Australia	2014	1747.937	43547.197	23596000	Oceania	54.280880
14	Australia	2015	1747.009	43923.258	23932000	Oceania	51.537594
15	Australia	2016	1734.215	44493.481	24263000	Oceania	54.122646

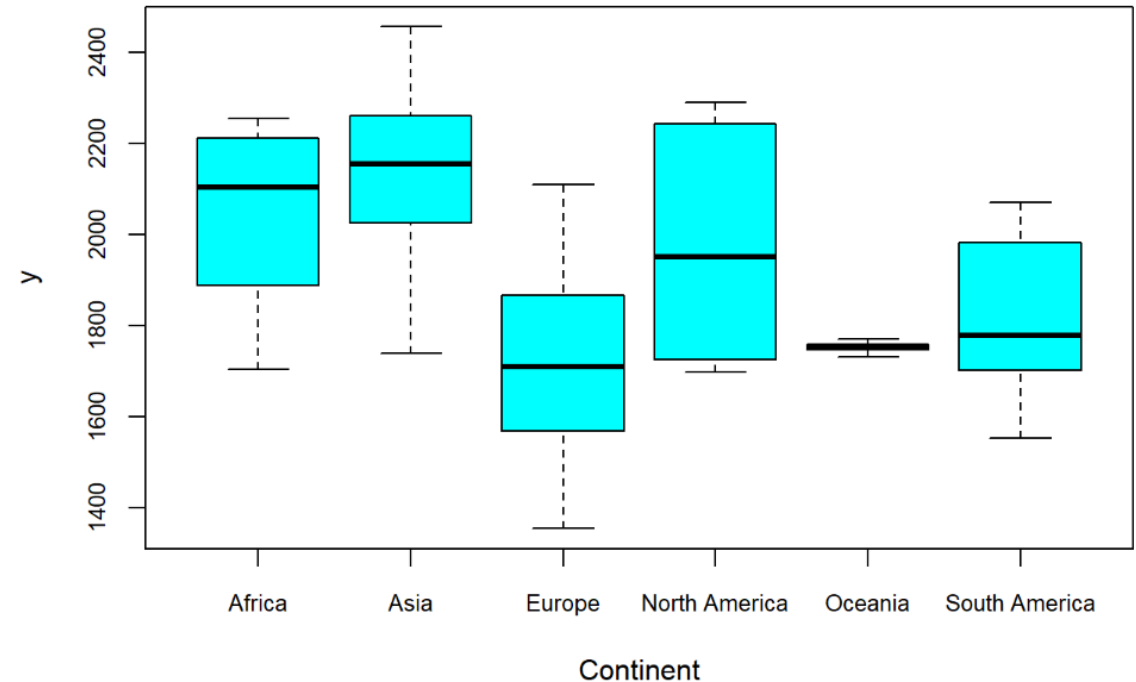
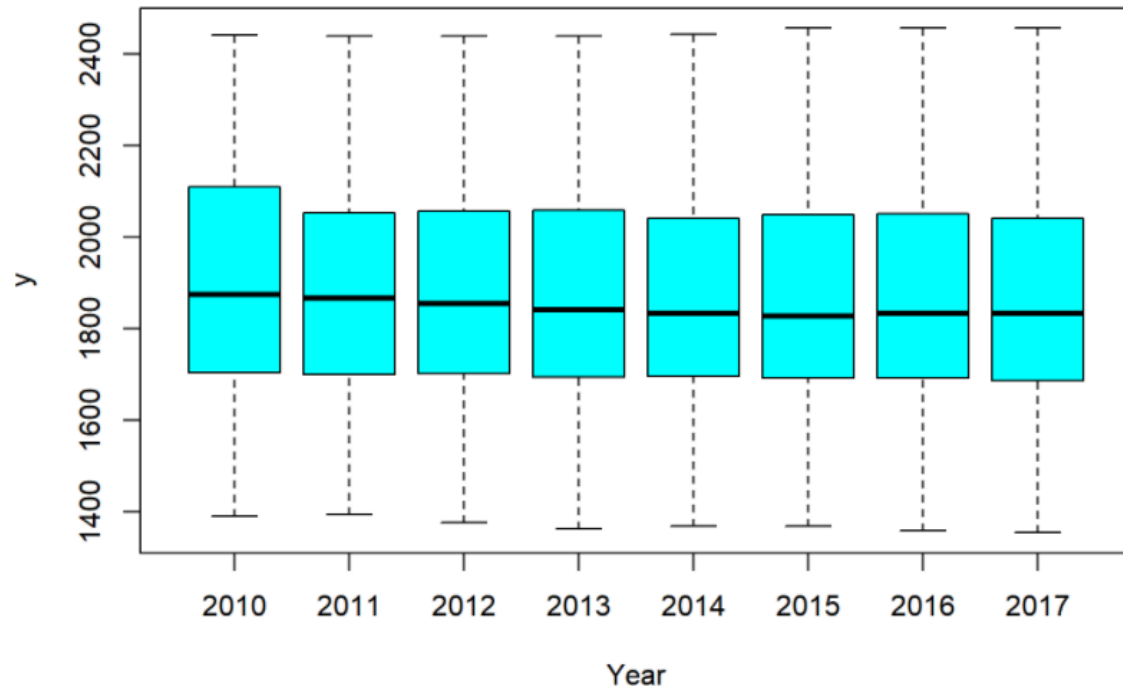
# Overview of data

---

```
##      Entity              Year  Average annual working hours      GDP
## Length:520      Min.    :2010      Min.    :1354              Min.    : 2443
## Class :character 1st Qu.:2012      1st Qu.:1692              1st Qu.:14661
## Mode  :character Median :2014      Median :1850              Median :25820
##                      Mean  :2014      Mean   :1873              Mean   :28411
##                      3rd Qu.:2015      3rd Qu.:2060              3rd Qu.:39777
##                      Max.    :2017      Max.    :2456              Max.    :94921
##      Population      Continent      Productivity
## Min.    :3.200e+05 Length:520      Min.    : 1.732
## 1st Qu.:5.494e+06 Class :character 1st Qu.: 16.634
## Median :1.767e+07 Mode  :character Median : 29.551
## Mean   :8.757e+07                      Mean   : 32.849
## 3rd Qu.:6.597e+07                      3rd Qu.: 48.304
## Max.   :1.421e+09                      Max.   :109.772
```

# Overview of data

- There are 5 independent variables in the data finally processed, among which GDP, productivity and population are continuous variables, while continent and Year are category variables.
- Use boxplot to see the distribution of category variables.



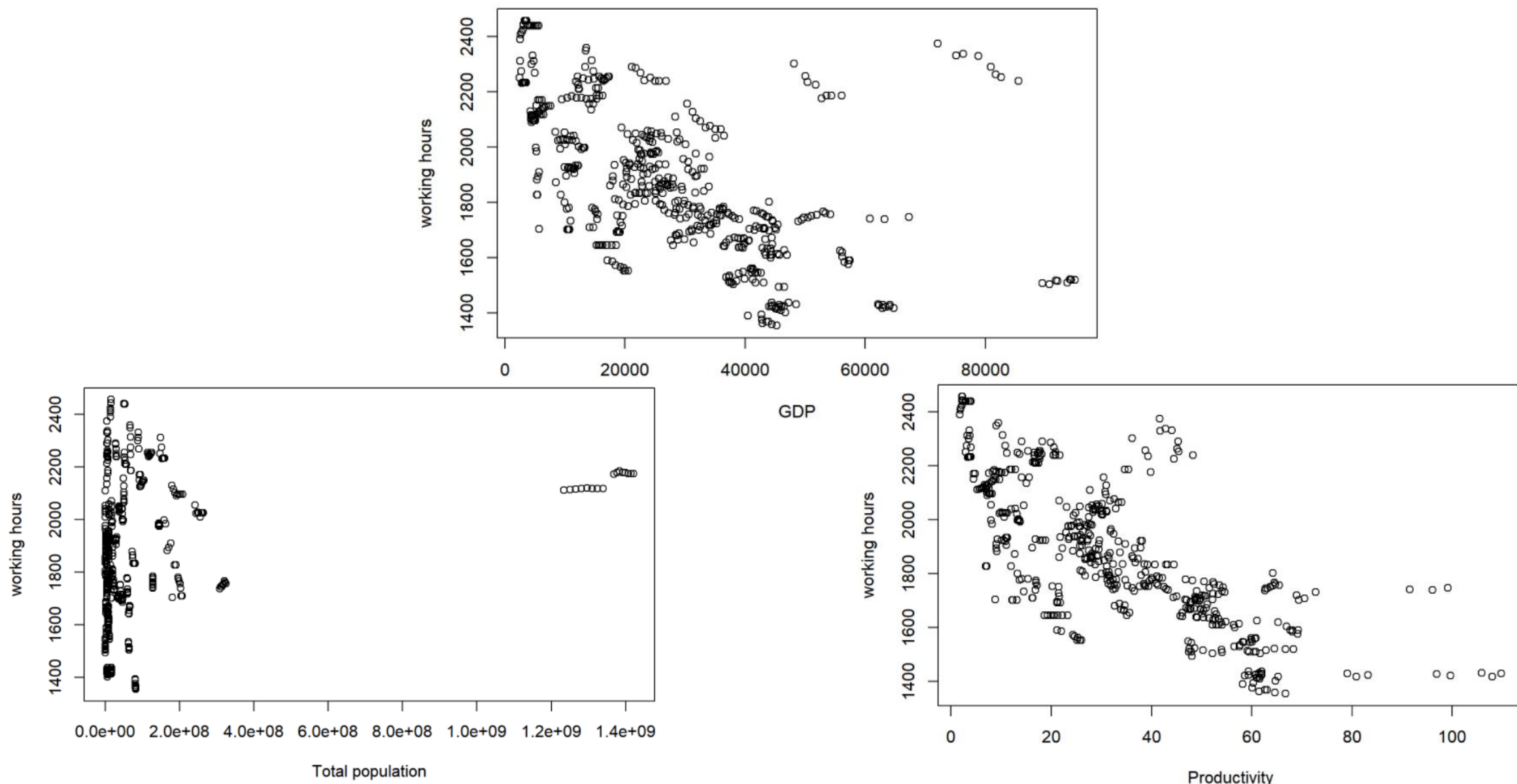


# Overview of data

---

- It can be seen from the figure that Year has little influence on working hours, so we can regard these data as the data of the same time period and delete the variable Year.
- The distribution of working hours in continent variable is significantly different, so we need to take it into account when building the model in the future.
- And then we check the distribution of continuity variables using a scatter plot.

# Overview of data



# **/03 Data analysis and Model**



# Data analysis and Model

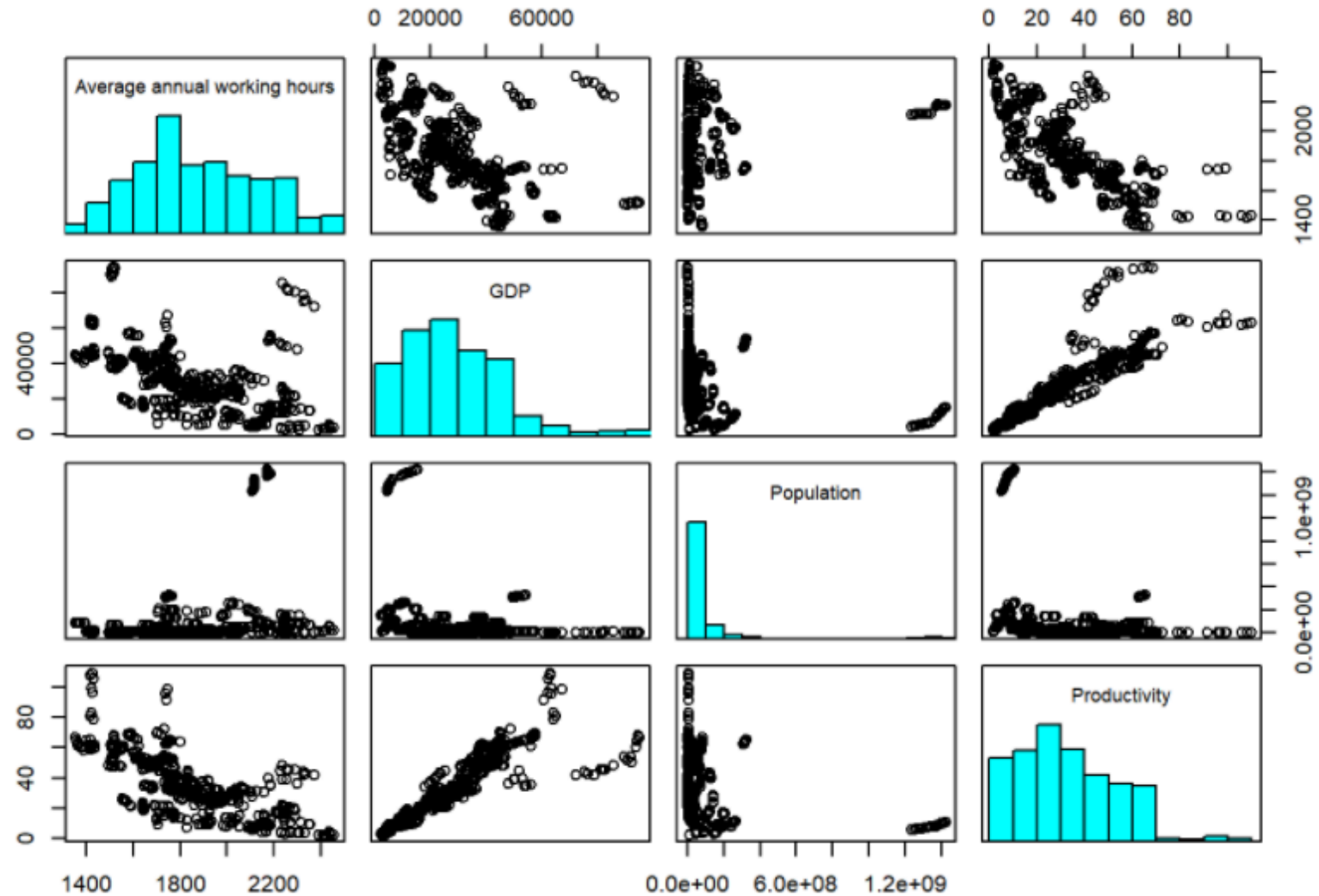
---

- First of all, we calculate the correlation coefficient of continuity variables to see which factor is related to working hours

```
##                               Average annual working hours   GDP Population
## Average annual working hours                1.00 -0.54      0.23
## GDP                                         -0.54  1.00      -0.26
## Population                                0.23 -0.26      1.00
## Productivity                             -0.73  0.86     -0.27

##                               Productivity
## Average annual working hours            -0.73
## GDP                                     0.86
## Population                             -0.27
## Productivity                           1.00
```

# Data analysis and Model



# Data analysis and Model

---

- It can be seen from the correlation coefficient that working hours is negatively correlated with GDP and Productivity, which is higher than GDP. There is a positive correlation between working hours and population, but the correlation is small. Among them, the correlation coefficient between GDP and Productivity is large, which is 0.86, indicating that the variable has multicollinearity.

# Data analysis and Model

- Building linear model---Parameter estimation for all variables

```
Call:
lm(formula = y ~ GDP + Productivity + Population + Year + Continent)
```

Residuals:

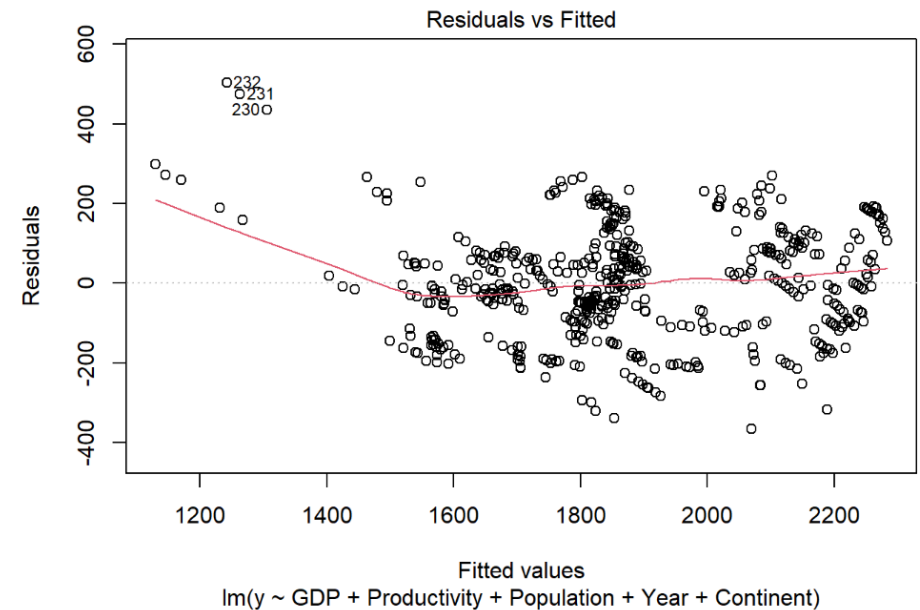
Min	1Q	Median	3Q	Max
-366.50	-95.81	-4.34	83.45	503.55

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	7.710e+03	5.292e+03	1.457	0.145759	
GDP	3.387e-03	6.697e-04	5.057	5.94e-07	***
Productivity	-1.046e+01	6.443e-01	-16.228	< 2e-16	***
Population	-8.966e-08	2.810e-08	-3.191	0.001507	**
Year	-2.755e+00	2.628e+00	-1.048	0.295109	
ContinentAsia	1.213e+02	3.660e+01	3.313	0.000989	***
ContinentEurope	-1.028e+02	3.726e+01	-2.758	0.006027	**
ContinentNorth America	9.816e+01	4.295e+01	2.286	0.022686	*
ContinentOceania	-6.850e+01	5.001e+01	-1.370	0.171381	
ContinentSouth America	-2.095e+02	3.893e+01	-5.380	1.14e-07	***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 136.9 on 510 degrees of freedom  
Multiple R-squared: 0.7332, Adjusted R-squared: 0.7285  
F-statistic: 155.8 on 9 and 510 DF, p-value: < 2.2e-16



# Data analysis and Model

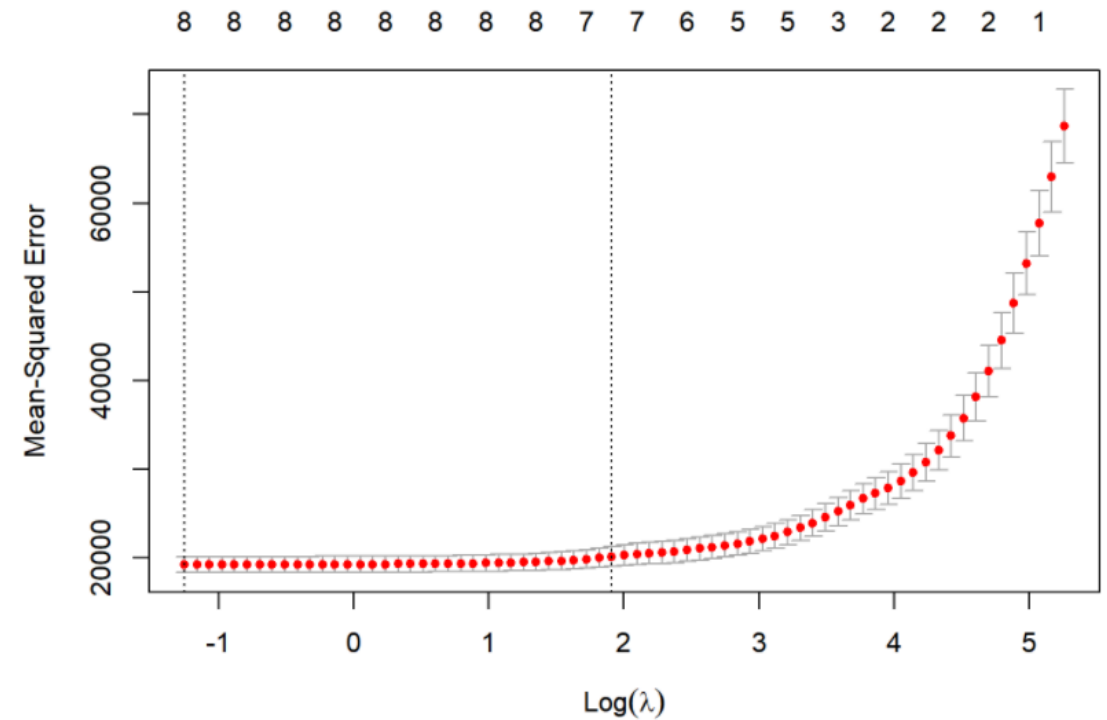
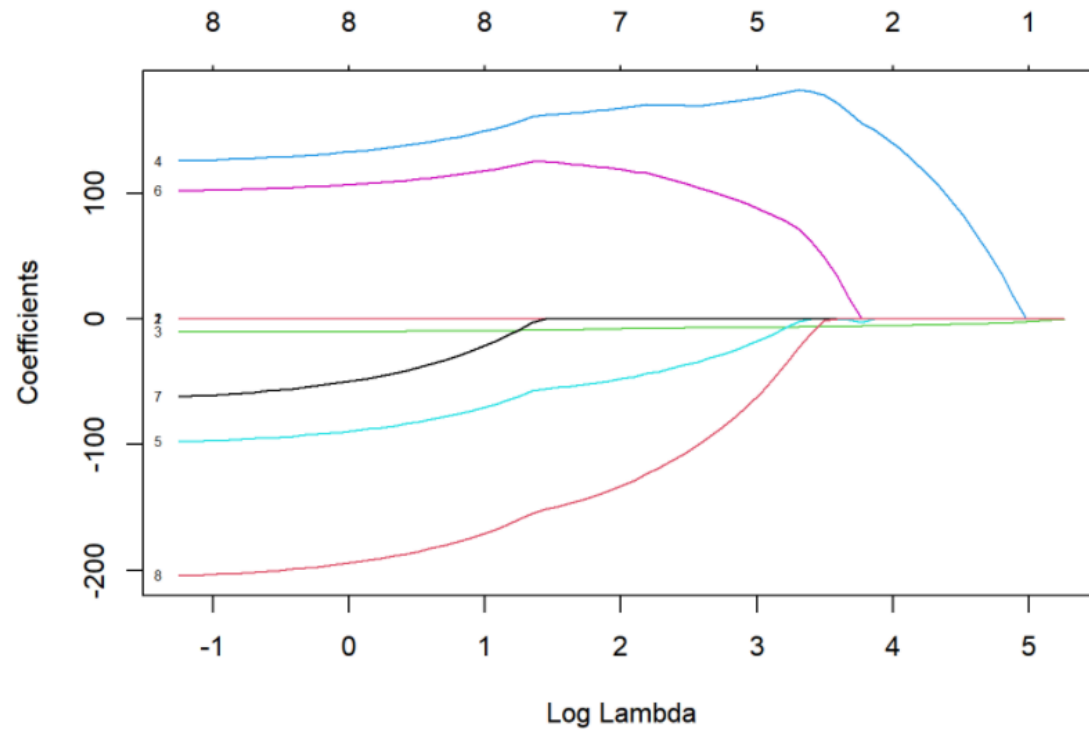
---

- According to the LM model that the p-value of year is large, which is consistent with the previous analysis, so this variable is considered to be deleted. In the correlation coefficient figure, it can be seen that GDP and Productivity are negatively correlated with working hours respectively. However, in the regression, the coefficient of GDP is positive, indicating that the model is incorrect.



# Data analysis and Model

- Exploit Lasso regression to deal with multicollinearity.



# Data analysis and Model

---

- Take the best value of lambda to look at the modeling situation, which balances the trade-off with variance and bias
- Compute the  $R^2$  of lasso model and we get 0.7868817

```
## 9 x 1 sparse Matrix of class "dgCMatrix"  
##  
## (Intercept)      2159.112  
## GDP              0.003  
## Population       0.000  
## Productivity     -10.366  
## dAsia            126.003  
## dEurope          -97.157  
## dNorth.America   102.260  
## dOceania         -61.329  
## dSouth.America  -203.879
```

# Data analysis and Model

---

- According to the variable selection by lasso regression, we can also reduce Year and Population variable in a new lm model to compared with the last one. Because the simpler the model, the stronger the generalization ability.

```
##
## Call:
## lm(formula = `Average annual working hours` ~ GDP + Productivity +
##     Continent)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -376.15  -92.42   -8.44   88.88  491.25
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.152e+03  3.490e+01  61.664 < 2e-16 ***
## GDP            3.594e-03  6.716e-04   5.352 1.31e-07 ***
## Productivity   -1.050e+01  6.501e-01 -16.150 < 2e-16 ***
## ContinentAsia    1.098e+02  3.675e+01   2.989 0.00293 **
## ContinentEurope -9.871e+01  3.758e+01  -2.627 0.00888 **
## ContinentNorth America  9.406e+01  4.331e+01   2.172 0.03031 *
## ContinentOceania  -6.431e+01  5.045e+01  -1.275 0.20298
## ContinentSouth America -2.050e+02  3.927e+01 -5.220 2.60e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 138.2 on 512 degrees of freedom
## Multiple R-squared:  0.7273, Adjusted R-squared:  0.7235
## F-statistic: 195.1 on 7 and 512 DF,  p-value: < 2.2e-16
```

# Data analysis and Model

---

- After regression, It can be seen that the points in the residual graph are randomly distributed on the horizontal axis, indicating that the linear regression model is reasonable. In the model, the Adjusted R-squared is 0.7087, and the points on the QQ diagram are approximately near a straight line, indicating that the sample data is approximately normal distribution.

# **/04 Conclusion**



# Conclusion

---

- Comparing two models with independent variables GDP, Productivity and Continent, the adjusted  $R^2$  of lasso is greater than the one of general linear model. So we choose lasso as our final model.

# Conclusion

---

- In our model, Average annual working hours is greatly affected by Productivity and Continent. The effect of GDP on working hours can be ignored since its coefficient is too small. From the coefficients, we can know that when other variables keep same, if Productivity increases and changes by one unit, Average annual working hours will decrease by about 10 hours. From the categorical variable Continent, we can see that, when another variables keep same, compared with Africa, the difference of Average annual working hours in Asia, Europe, North America, Oceania, and South America are about 126, -97, 102, -61, and -204 hours, respectively.
- So the data tells us that if we increase our productivity, we may spend less time on working. Meanwhile, it tells us that the continent which the country belongs to is also a factor of analyzing working time. Compared with Africa, only Asia and North America have longer working time.

# Conclusion

---

- What can we learn from the data?
- Work takes up most of people's time these days, and in countries with more advanced economies and higher productivity, such as Europe, people work fewer hours. In Asia and North America, however, people work more hours even though their GDP and productivity are not low, reflecting a greater emphasis on work and a higher proportion of work in people's lives.
- Why is that?
- We think the reason may be that in Asia and North America, people are under more competitive pressure.



# Conclusion

---

- Work is just a part of life. We don't want it to take up too much of your life. Hope that everyone can keep a balance between work and life.

# THANKS

---

