

# **CEO Characteristics and Firm R&D Spending**

Math Capstone PBL (Data Analysis) - Project #1

---

Jaeseon Lee<sup>1</sup> Junwoo Yang<sup>2</sup>

October 27, 2020

<sup>1</sup>Department of Economics & Finance  
Hanyang University

<sup>2</sup>Department of Finance  
Hanyang University

# Outline

1. Introduction
2. EDA
3. Correlation Analysis
4. Conclusion

# Introduction

---

### **Does the CEO's nature affect the company's R&D investment?**

- What financial variables are relevant to the extent of the firm's R&D investment?
- Does the variable that characterizes the CEO lack explanatory power for R&D investments?
- If so, what characteristics correlate?

## Dimension of raw data

**1726 rows:** U.S. S&P 1,500 + other Execucomp-reporting firms

**27 columns:** financial variables + CEO characteristics

## Data source

Wharton Research Data Services (WRDS) Compustat Data / Execucomp<sup>1 2</sup>

<sup>1</sup><http://wrds-www.wharton.upenn.edu.ssl.access.hanyang.ac.kr/>

<sup>2</sup><https://lib.hanyang.ac.kr/#/er/web>

## Variables for identifying firms

**gvkey:** A firm's identifier used by S&P Capital IQ / Compustat / Execucomp

**sic:** A four-digit number that identifies a firm's primary industry of operation <sup>3</sup>

**state:** A two-letter code denoting the state in which the firm is headquartered

---

<sup>3</sup>Variables SIC2D and SIC3D denote the first two or three letters of the SIC code - the first two letters denote the broad classification, the third letter the middle classification

## Financial variables (1)

**size:** firm size (total assets)

**bm:** book-to-market

**fcf:** free cash flow (divided by total assets)

**hhil:** Herfindhal-Hirschman Index <sup>4</sup>

**opperf:** operating performance (divided by total assets)

**leverage:** market value of a firm's leverage (divided by total mark-to-market assets)

**tobinsq:** Tobin's Q

**rndmissing:** a dummy variable that equals 1 if a firm's R&D expenditure is missing in the financial statement

---

<sup>4</sup>The more competitive the industry a firm operates in, the lower this number. It is always between 0 and 1.

## Financial variables (2)

**rndratio:** a firm's R&D expenditure (divided by total assets)

**roa:** return on assets (divided by total assets)

**salesgrowth:** year on year growth rate of a firm's sales

**cashratio:** a firm's cash holdings (divided by total assets)

**divpay:** a dummy variable that equals 1 if and only if a firm has paid out dividends in the same fiscal year.

**intan:** intangibility measure. The higher this number, the more intangible a firm's assets<sup>5</sup>

**invest:** a firm's investment ratio<sup>6</sup>

---

<sup>5</sup>(total assets – property, plants, and equipments value)/total assets

<sup>6</sup>investment/(property, plants, and equipments value)

## CEO characteristics

ceoage: CEO's age

ceocomp: CEO's total compensation, in \$1000

insiderceo: CEO who was promoted from inside the company (as opposed to an outsider)

femaleceo: a dummy variable that equals 1 if the CEO is female

ceopayslice: CEO's compensation divided by the firm's top 5 earning directors' total compensation <sup>7</sup>

ceoequity: CEO's holdings of the firm's stock

ceotenure: CEO's tenure in his or her current position, in years

---

<sup>7</sup>The higher this number, the more disproportionately highly a CEO is compensated within the top management team. Bebchuk, Cremers, and Peyer (2011) argue this is a good measure of how powerful a CEO is within the company.

**EDA**

---

# Handling missing values

	#	%
gvkey	0	0
sic	0	0
sic2d	0	0
sic3d	0	0
state	56	3.24
size	1	0.06
bm	18	1.04
fcf	104	6.03
hhi	0	0
upperf	104	6.03
leverage	36	2.09
tobinsq	18	1.04
rndmissing	0	0
rndratio	779	45.13
roa	2	0.12
salesgrowth	5	0.29
divpay	0	0
cashratio	2	0.12
intan	101	5.85
invest	120	6.95
ceoage	1	0.06
ceocomp	1	0.06
insiderceo	0	0
femaleceo	0	0
ceopayslice	266	15.41
ceoequity	29	1.68
ceotenure	21	1.22

Mean groupby sic

**3672:** Printed Circuit Boards



Mean groupby sic3d  
(middle classification)

**367:** Electronic, Component and  
Accessories

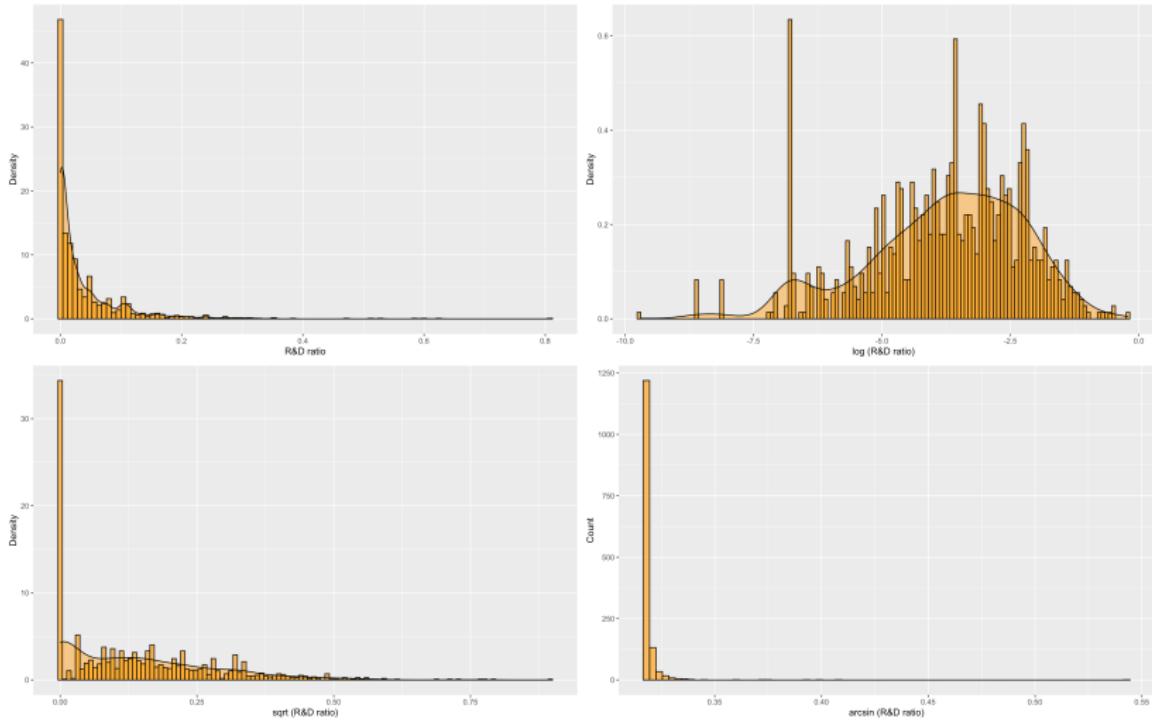


Mean groupby sic2d  
(broad classification)

**36:** Electronic and Other  
Equipment

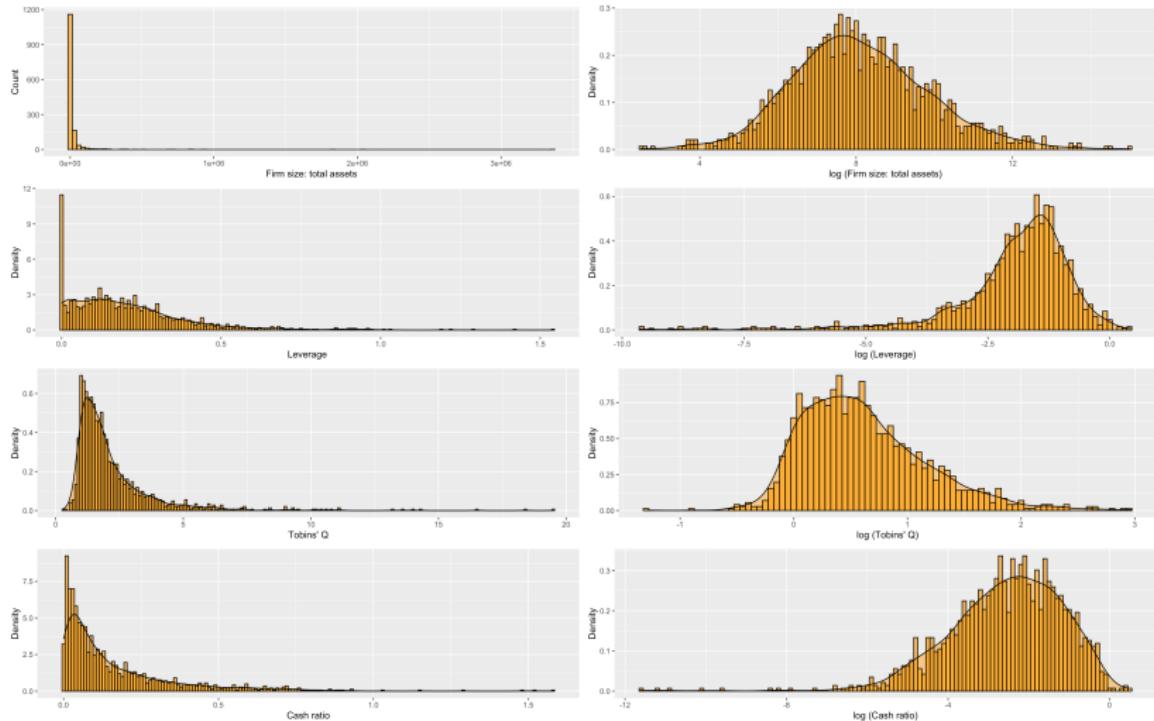
**Table 1:** Missing values of raw data

# Data transformation



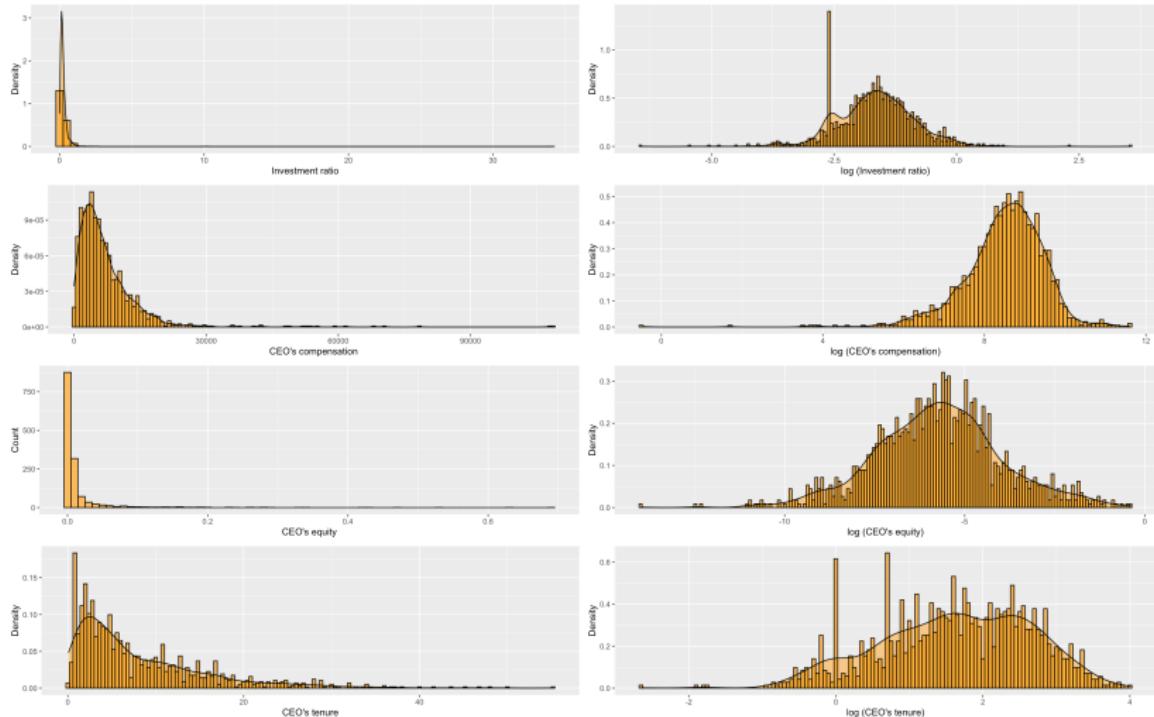
**Figure 1:** Transformations of rndratio

# Data transformation



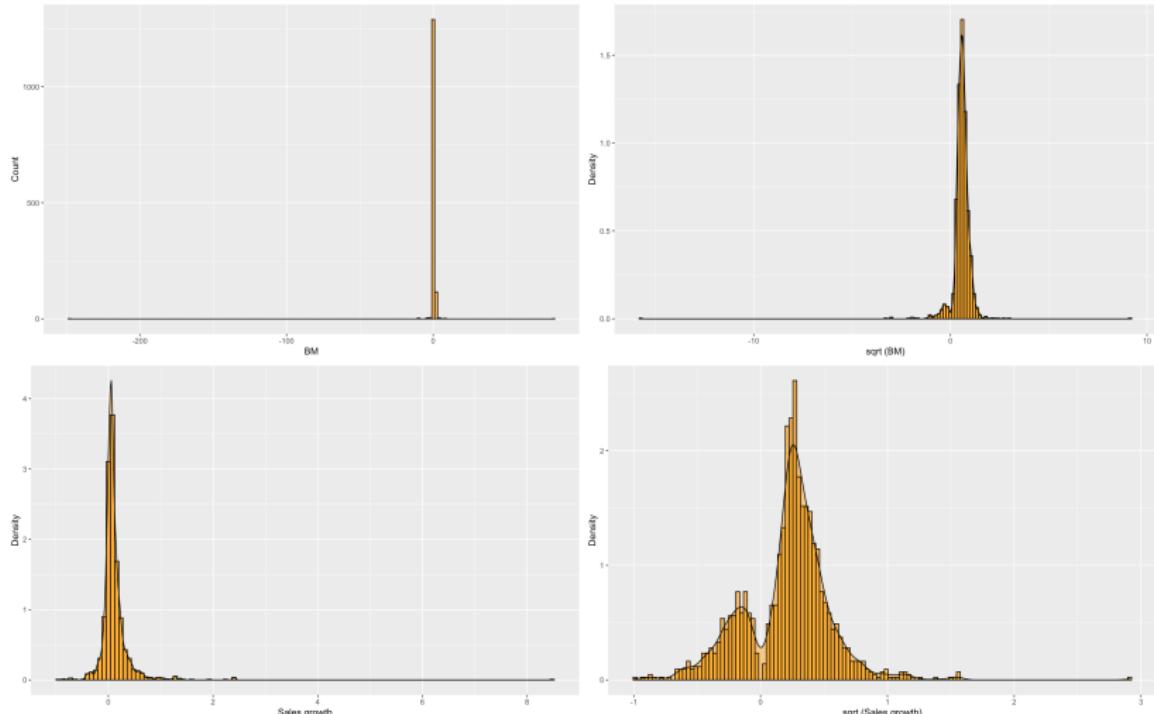
**Figure 2:** Log transformations of size, leverage, tobinsq, cashratio

# Data transformation



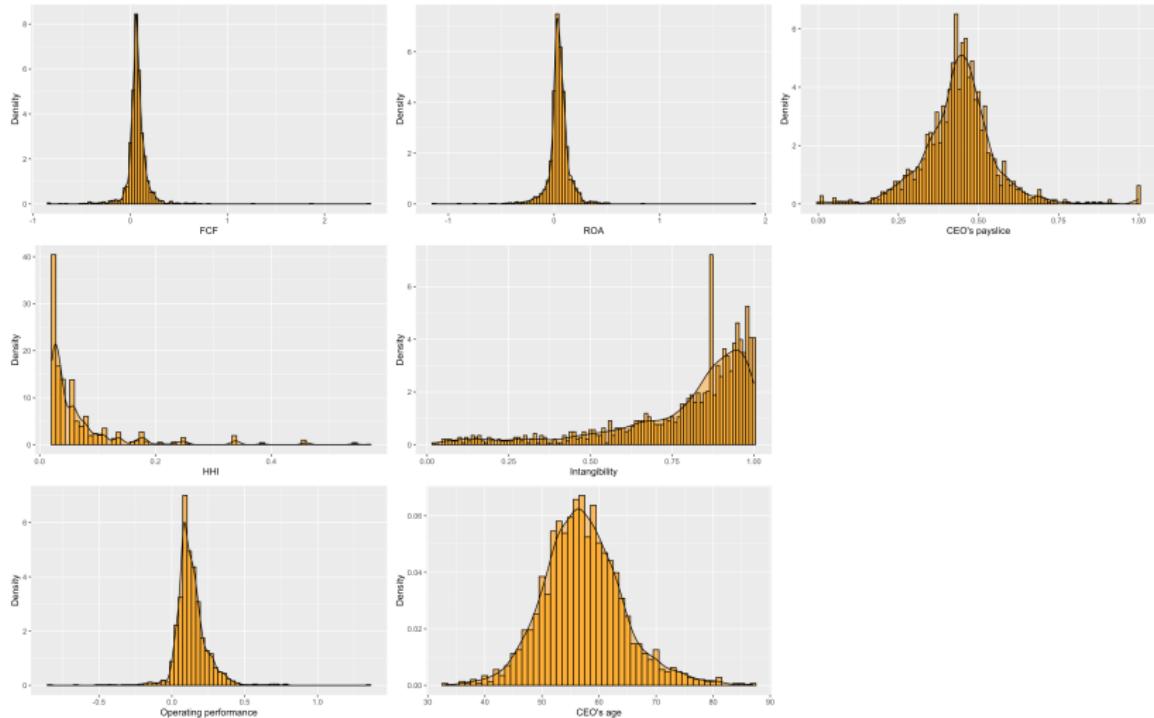
**Figure 3:** Log transformation of invest, ceocomp, ceoequity, ceotenure

# Data transformation



**Figure 4:** Square-root transformation (with sign) of  $bm$ ,  $salesgrowth$

# Data transformation



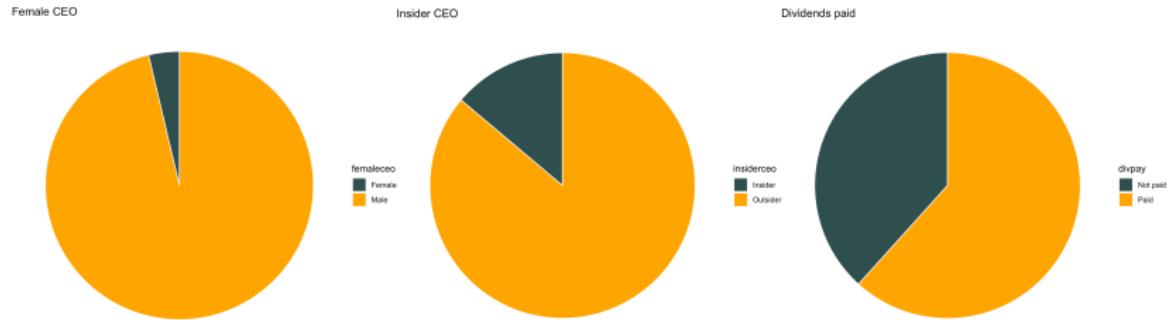
**Figure 5:** Not transformed

## Summary statistics

	mean	sd	Q1	Q3	max	min	kurtosis	skewness
log size	8.098	1.745	6.941	9.163	15.023	2.483	0.470	0.333
log leverage	-1.951	1.266	-2.319	-1.217	0.435	-9.648	8.524	-2.336
log tobinsq	0.613	0.544	0.212	0.914	2.973	-1.281	1.105	0.869
log cashratio	-2.568	1.421	-3.418	-1.543	0.458	-11.645	3.454	-1.068
log invest	-1.649	0.798	-2.129	-1.171	3.530	-6.430	3.174	-0.060
$\sqrt{bm}$	0.569	0.643	0.435	0.765	9.096	-15.762	313.873	-10.471
$\sqrt{salesgrowth}$	0.213	0.339	0.068	0.391	2.910	-1.000	3.755	0.187
fcf	0.073	0.140	0.329	0.106	2.447	-0.844	85.110	4.820
hh	0.068	0.077	0.026	0.073	0.566	0.021	12.951	3.296
opperf	0.133	0.120	0.080	0.178	1.363	-0.832	16.421	0.656
roa	0.045	0.120	0.010	0.087	1.878	-1.128	50.972	1.169
intan	0.796	0.212	0.722	0.946	1.000	0.015	2.230	-1.631
rndratio	0.039	0.069	0.000	0.050	0.809	0.000	25.756	4.033
ceoage	33.000	7.060	53.000	61.000	87.000	33.000	0.996	0.407
log ceocomp	8.489	1.003	7.994	9.135	11.596	-0.453	7.256	-1.380
log ceoequity	-5.725	1.774	-6.858	-4.699	-0.377	-13.999	0.671	-0.061
log ceotenure	1.625	1.023	0.916	2.418	4.009	-2.643	-0.339	-0.328
ceopayslice	0.440	0.116	0.381	0.494	1.000	0.001	4.312	0.520

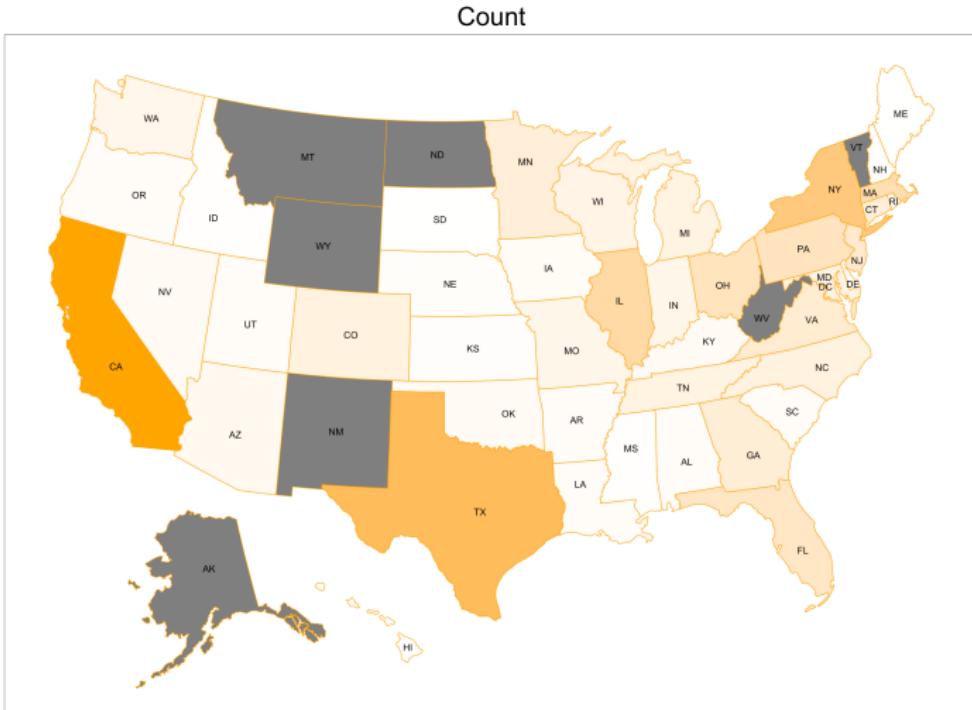
**Table 2:** Summary statistics for continuous variables

# Summary statistics



**Figure 6:** Pie chart of femaleceo, insiderceo, divpay

## Summary statistics



**Figure 7:** Count by state

## Correlation Analysis

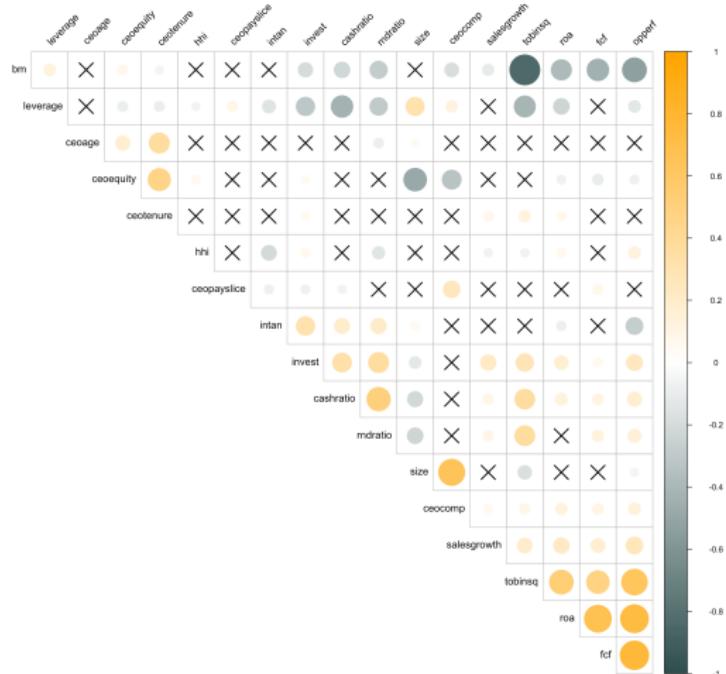
---

## rndratio — continuous variable

	$\tau$	p-value
roa	0.025	0.174
log ceocomp	-0.004	0.842
log ceoquity	0.012	0.493
log ceotenure	0.028	0.122
<u>ceopayslice</u>	-0.027	0.144

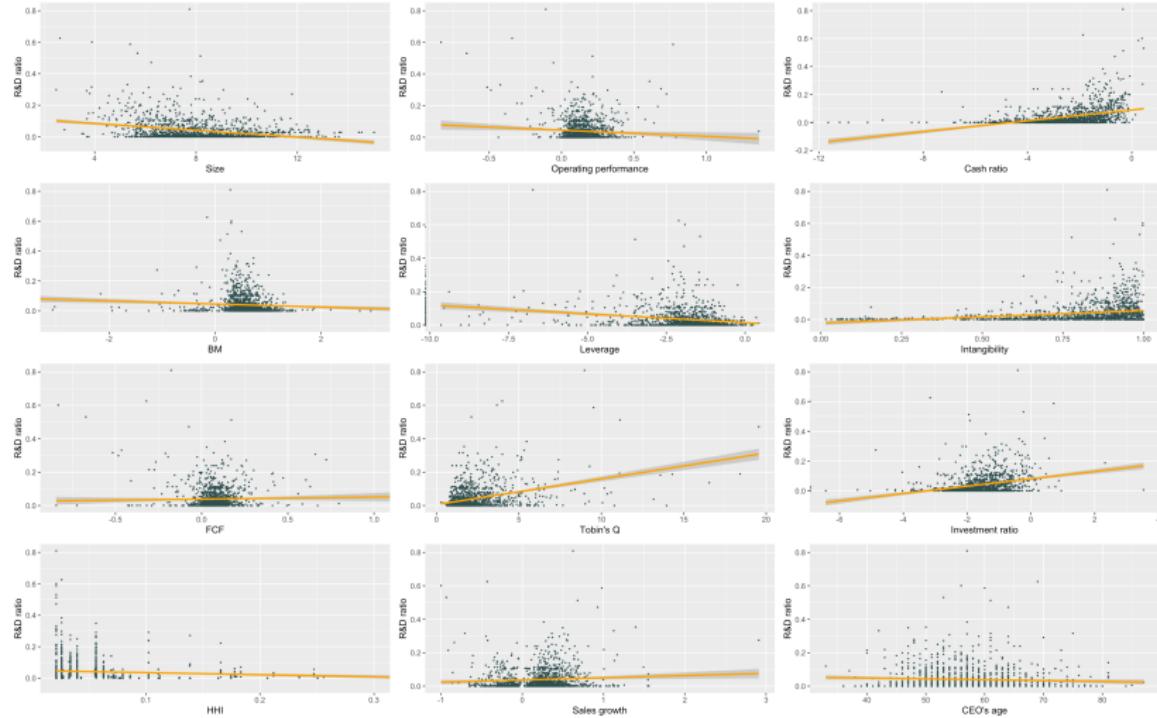
**Table 3:** Kendall correlation test of variables that can't reject  $H_0$

# rndratio — continuous variable



**Figure 8:** Combining Spearman correlogram with significance test

# rndratio — continuous variable



**Figure 9:** Scatter plot and regression line

# rndratio — continuous variable

	Dependent variable: rndratio					
	(1)	(2)	(3)	(4)	(5)	(6)
log size	-0.011*** (0.001)					
log leverage		-0.011*** (0.001)				
log tobinsq			0.015*** (0.001)			
log cashratio				0.019*** (0.001)		
log invest					0.025*** (0.002)	
ceoage						-0.001** (0.0003)
Constant	0.127*** (0.008)	0.014*** (0.003)	0.006** (0.003)	0.089*** (0.003)	0.081*** (0.004)	0.069*** (0.015)
Observations	1,429	1,301	1,429	1,426	1,402	1,429
R <sup>2</sup>	0.074	0.043	0.147	0.157	0.084	0.003
Adjusted R <sup>2</sup>	0.074	0.043	0.147	0.157	0.083	0.002
Residual Std. Error	0.067 (df = 1427)	0.063 (df = 1299)	0.064 (df = 1427)	0.064 (df = 1424)	0.065 (df = 1400)	0.069 (df = 1427)
F Statistic	114.643*** (df = 1; 1427)	59.035*** (df = 1; 1299)	246.718*** (df = 1; 1427)	266.045*** (df = 1; 1424)	128.159*** (df = 1; 1400)	3.897** (df = 1; 1427)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 4:** Stargazer for simple linear regression (1)

# rndratio — continuous variable

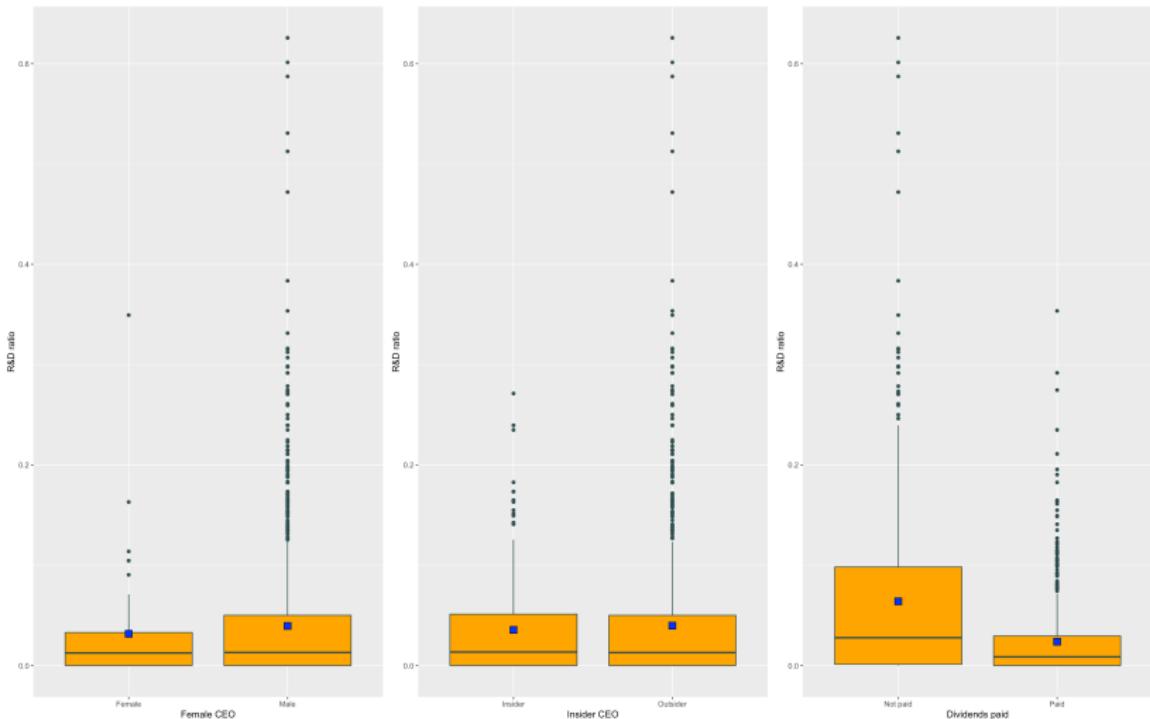
	Dependent variable: rndratio					
	(1)	(2)	(3)	(4)	(5)	(6)
$\sqrt{bm}$		-0.010*** (0.003)				
$\sqrt{salesgrowth}$			0.013** (0.005)			
fcf				0.012 (0.013)		
hhci					-0.134*** (0.024)	
opperf						-0.039*** (0.015)
intan						0.078*** (0.008)
Constant	0.045*** (0.002)	0.037*** (0.002)	0.039*** (0.002)	0.049*** (0.002)	0.045*** (0.003)	-0.023*** (0.007)
Observations	1,429	1,429	1,429	1,429	1,429	1,429
R <sup>2</sup>	0.009	0.004	0.001	0.022	0.005	0.057
Adjusted R <sup>2</sup>	0.008	0.004	-0.0001	0.021	0.004	0.057
Residual Std. Error (df = 1427)	0.069	0.069	0.069	0.069	0.069	0.067
F Statistic (df = 1; 1427)	12.984***	6.020**	0.846	32.064***	6.706***	86.702***

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 5:** Stargazer for simple linear regression (2)

# rndratio — categorical variable



**Figure 10:** Box plot by femaleceo, insiderceo, divpay

**Welch's t-test (in which a two-tailed test is applied)**

$H_{T,0}$ : The two population means are equal.

$H_{T,1}$ : True difference in means is not equal to 0.

**F-test of equality of variances** <sup>8</sup>

$H_{F,0}$ : Two normal populations have the same variance.

$H_{F,1}$ : True ratio of variances is not equal to 1.

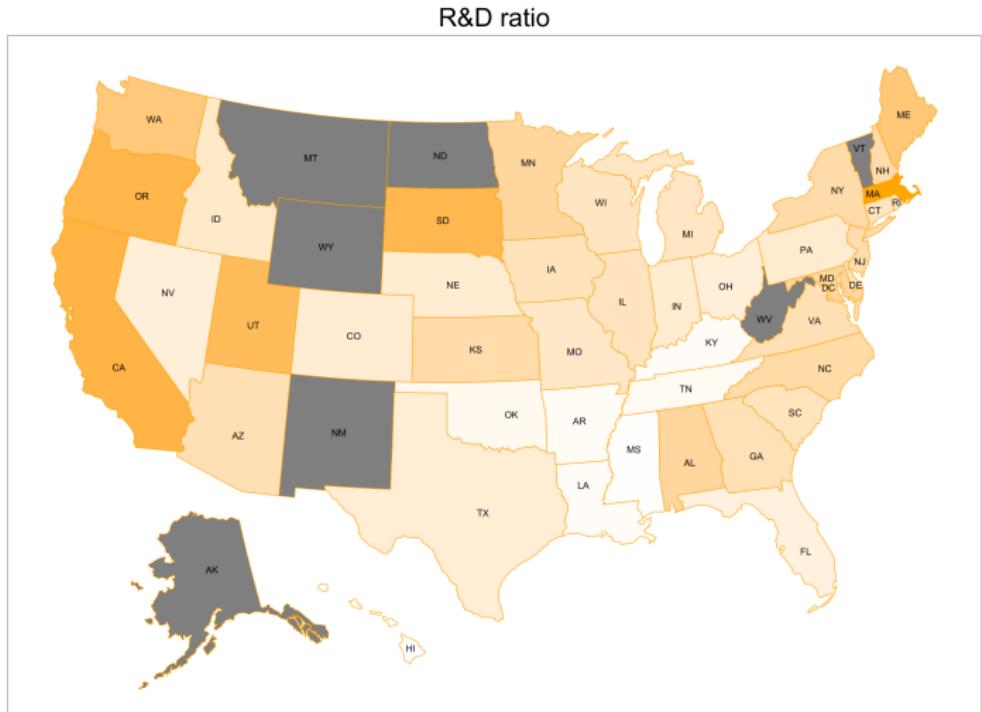
	$F$	$t$	accept
divpay	$< 2.2 \times 10^{-16}$	0.3033	$H_{F,1}, H_{T,0}$
insiderceo	$3.995 \times 10^{-09}$	0.244	$H_{F,1}, H_{T,0}$
femaleceo	0.06457	0.3895	$H_{F,0}, H_{T,0}$

**Table 6:**  $p$ -values and result at  $\alpha = 0.05$

---

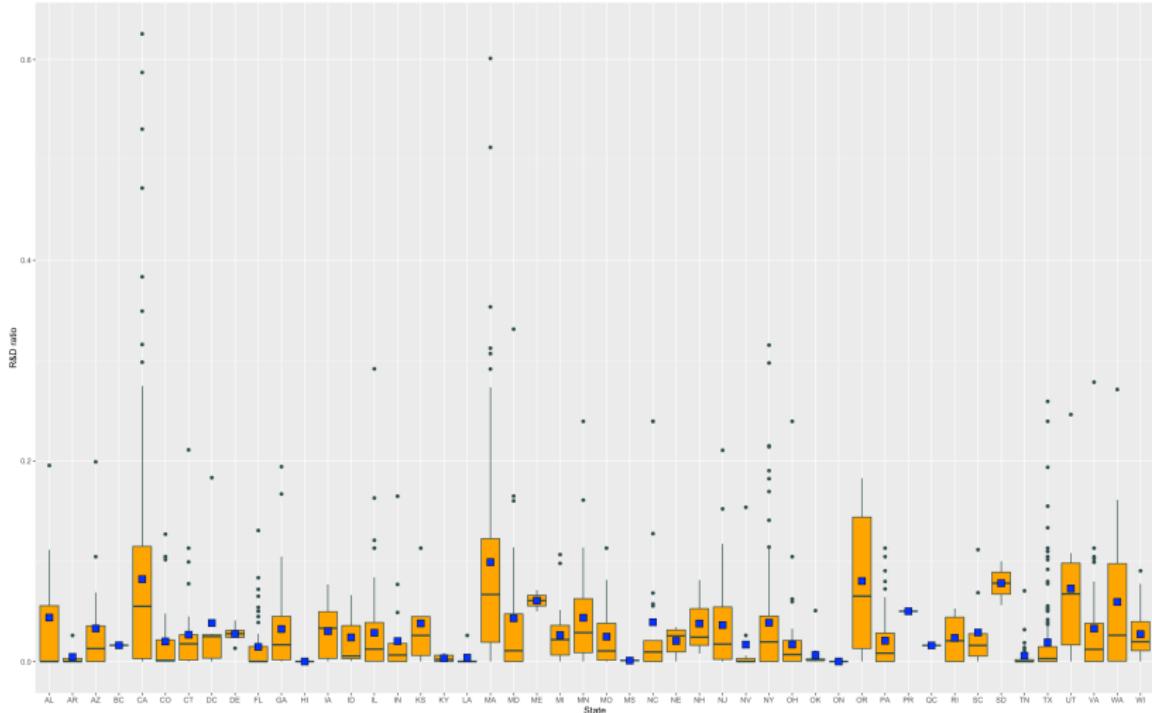
<sup>8</sup> $H_{F,0}$ : Homoscedasticity,  $H_{F,1}$ : Heteroscedasticity

# rndratio — categorical variable



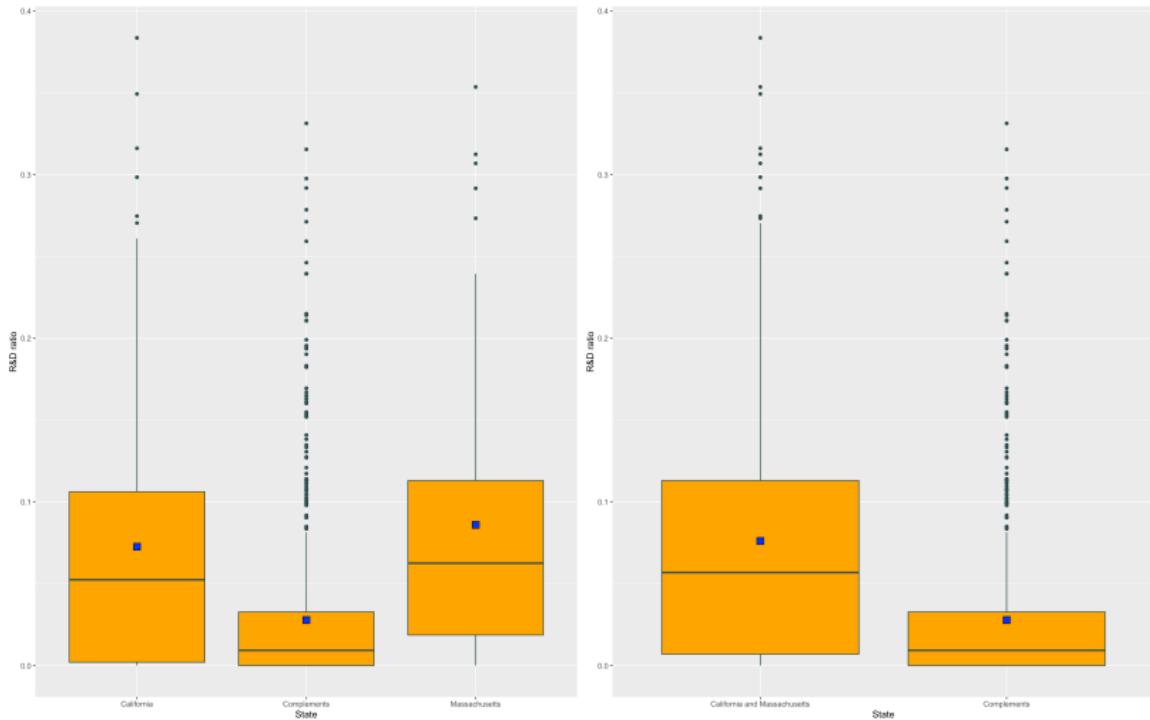
**Figure 11:** Mean of rndratio by state

## rndratio — categorical variable



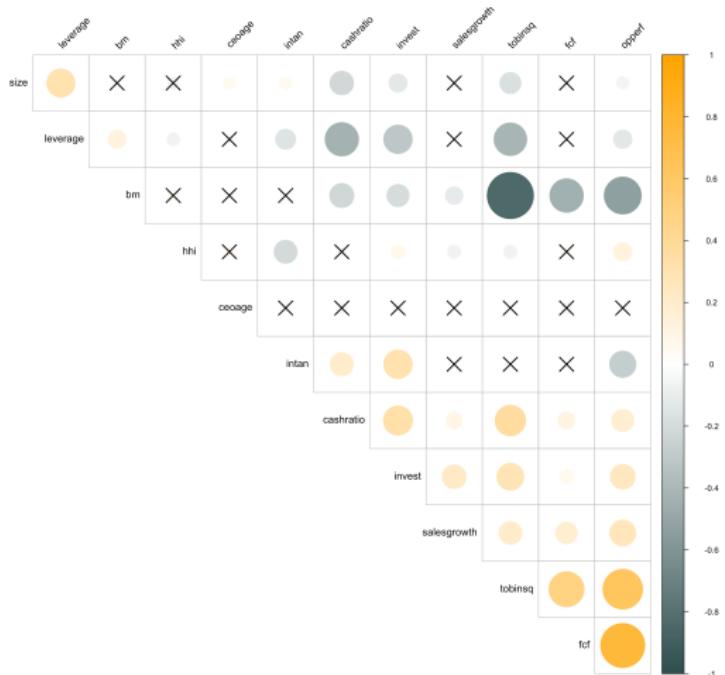
**Figure 12:** Box plot by state

## rndratio — categorical variable



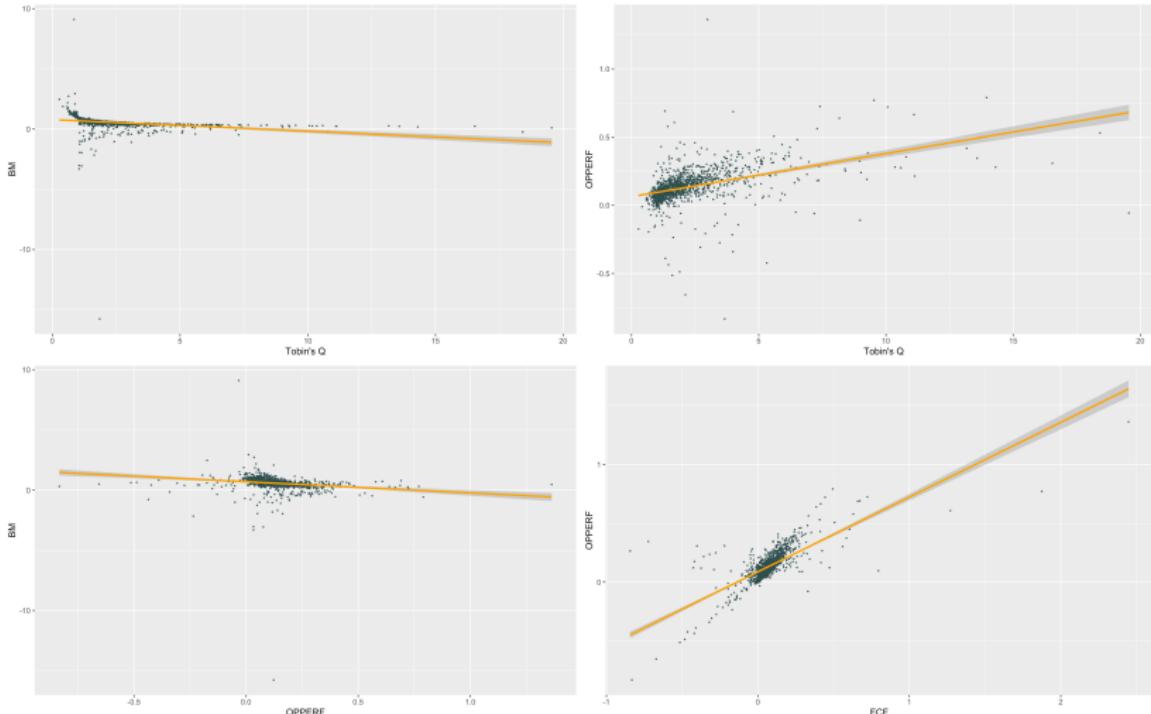
**Figure 13:** Regrouping state

# Between continuous variables



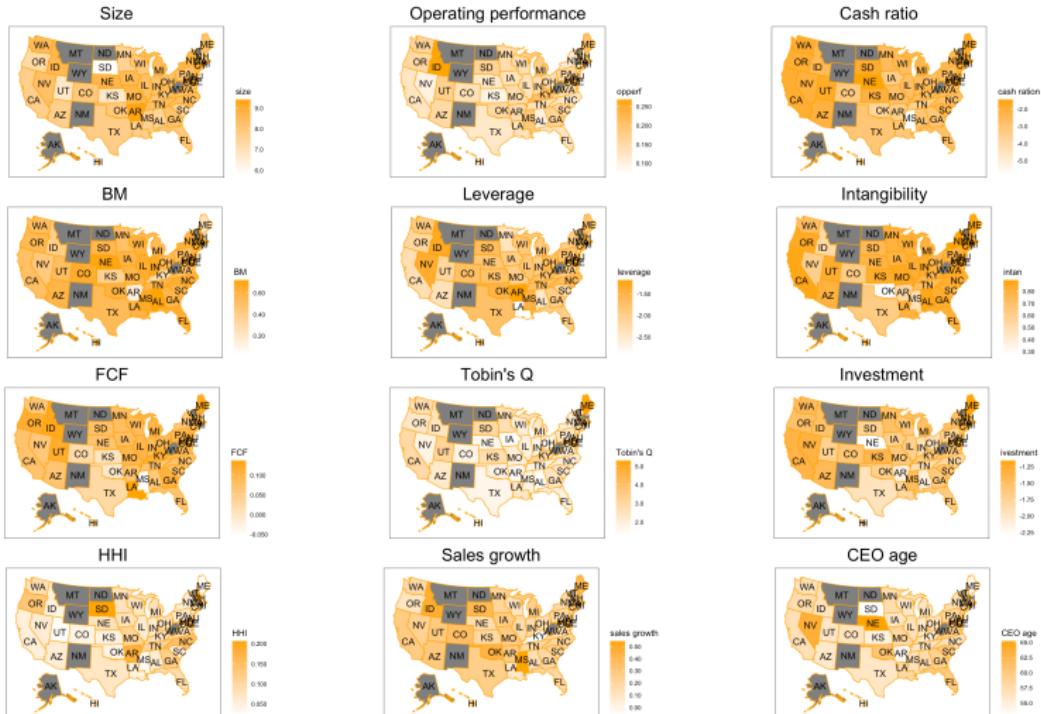
**Figure 14:** Combining Spearman correlogram with significance test

## Between continuous variables



**Figure 15:** Cases of  $|\rho| > 0.5$

# Between categorical and continuous variables



**Figure 16:** Surviving continuous variables by state (indistinguishable)

## Conclusion

---

## Answers to subquestions

### **What financial variables are relevant to the extent of the firm's R&D investment?**

All financial variables used except ROA.

### **Does the variable that characterizes the CEO lack explanatory power for R&D investments?**

Correlation analysis confirms that not all CEO characteristics are uncorrelated with R&D ratio.

### **If so, what characteristics correlate?**

CEO age is most relevant to the degree of R&D investment.

## Remaining analysis

- Check effects of normalization of R&D ratio (standard score)
- Eliminate multicollinearity based on VIF
- Select variables: FS, BS, SM
- Residual Analysis: Regression Assumptions, Influence and Outliers

## References i

-  Chatterjee, Benmelech, E., and Frydman, C. (2015)  
**Military CEOs.**  
*Journal of Financial Economics*, 117,(1), 43-59.
-  Barker III, V. L., and Mueller, G. C. (2002)  
**CEO characteristics and firm R&D spending.**  
*Management Science*, 48(6), 782-801.
-  Serfling, M. A. (2014)  
**CEO age and the riskiness of corporate policies.**  
*Journal of Corporate Finance*, 25, 251-273.

-  Kronmal, R. A. (1993)  
**Spurious correlation and the fallacy of the ratio standard revisited.**  
*Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 156(3), 379-392.
-  Kang H. (2013)  
**The prevention and handling of the missing data.**  
*Korean journal of anesthesiology*, 64(5), 402-406.
-  S. and Hadi, A.S. (2012)  
**Regression Analysis by Example.**  
Wiley, New York.