

# Plan: Big data research team

Sung Won Kang  
2016년 9월 19일

# What we do

- 우리는 환경관련 Data Journalism을 시작한다

# Data Journalism?

- Data Journalism = Empirical social science ON DEADLINE
  - 시의성 있는 주제를 선택한다
  - 정교하지만 해석은 가능한 분석도구를 이용하여 분석한다.
  - 주어진. 혹은 설정한 기간 내에 발신한다.
- 방법론은 정확하고, 신속하고, 투명하여야 한다.
  - 재생 가능한 연구 (reproducible research)를 지향한다.

<http://blog.revolutionanalytics.com/2016/07/data-journalism-with-r-at-538.html>

[https://sec.ch9.ms/ch9/3543/cb72d91a-95b3-4950-aafb-2317901e3543/D1McCAWS11Flowers\\_mid.mp4](https://sec.ch9.ms/ch9/3543/cb72d91a-95b3-4950-aafb-2317901e3543/D1McCAWS11Flowers_mid.mp4)

# 연구의 흐름

1. 주제 선정 : 4대 일간지 website webscraping + 자연언어분석
2. 자료 수집: 자료 수집 및 정리기능을 자동화
3. 환경관련분석 : 오염도 추정
4. 경제관련분석 : 일반균형 분석 및 부분균형 분석
5. 결론 및 시사점

# 발간방식

- 정기적 발신: Rmarkdown 이용 html 형태로 실시간 발신

```
##  
## Call:  
## lm(formula = y ~ x, data = d)  
##  
## Coefficients:  
## (Intercept)          x  
##    4.98489    0.03465
```

```
## (Intercept)          x  
## 4.98488812 0.03464878
```

- 보고서 발신: 연 1회 정기발신을 모아서 연구소 보고서 양식으로 발간

# 발간주기

- 2017년 : 연 2회 (3-6, 7-10) + 1회 (11-2018.2)
- 2018년 : 연 4회 (분기 당)
- 2019년 : 연 6회 (2개월 당) + 특별호 (부정기)
- 2020년 : 연 10회 (8월, 12월 제외) + 특별호 (부정기)

# 역할분담(잠정)

- 총괄: 장기복, 강성원
- 주제 선정: 추후 선정(채용?)
- 환경관련분석(교육): 이동현
- 경제관련분석(교육): 강성원
- 자료관리: 한국진
- 출판: 김진형
- General Assistant: 김진형
- 행정: 최두연

# 당면과제 (Due to Dec. 2016)

- 분석 tool 구축 (Donghyun,Sungwon): Deep Learning algorithm, General Eq, Partial Eq
- 자료 관리 과정 구체화 (Gookjin): source 파악, data scraping, storage, storage access.
- 출판 process 구체화 (Jinhyung): homepage 개설, real time publication (desktop -> website), data 및 code publication.



# 부탁의 말씀

- slack 'Data Is Art' team 상시 접속 유지
- 일상적인 요청 및 제출은 Slack을 이용 (공문은 intranet 계속 이용)
- '당면과제' 관련 Action plan 마련: Slack에 post (due to 09/31)
- 아는 것 나누기
  - 예: <http://blog.revolutionanalytics.com/>