





주제 선정 이유_공약과 정책



국공립어린이집 이용이동 이용이동 %까지확대

- · 초등돌봄교실 전학년 확대 등 초등생 안전돌봄책임제
- · 아빠육아휴직보너스제
- · 초등학교 입학 전 자녀를 둔 엄마, 아빠에게 임금감소 없는 근로시간 단축과 유연근무제 도입
- · 전업주부 등 고용보험 미가입 여성에게도 3개월간 월 50만원씩 출산수당 지급
- · 육아휴직급여 인상 (최초 3개월간 80%, 4개월 차부터는 50%로 인상)
- · 경력단절여성 취업알선책임제와 채용장려금 1년간 지원

문재인 대통령 "올해 국공립어린이집 450개 만들겠다"

○ 이성교 │ ② 승인 2018.01.24 16:47

뉴스홈 > 정치

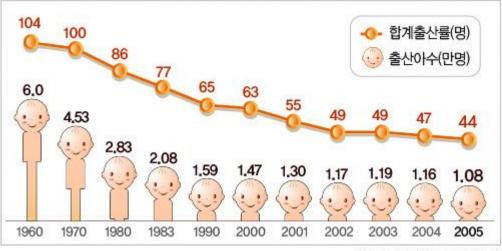
문재인 대통령 "임기 내 국공립 어린이집 40%로 확충 가능"

주제 선정 이유_국공립 어린이집 부족





주제 선정 이유_최저 출산율

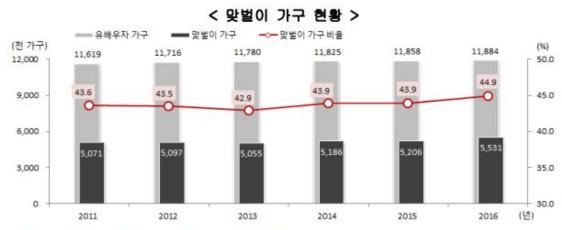


자료: 통계청, '인구동태통계연보'

주제 선정 이유_의문



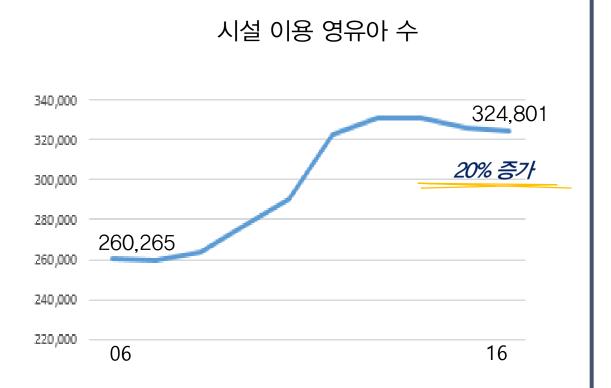
주제 선정 이유_문제의 원인

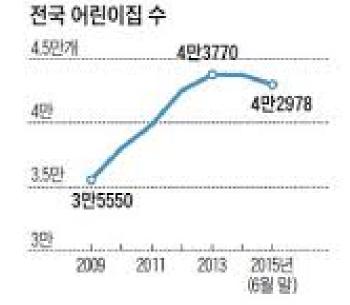


자료 : 통계청,「지역별고용조사」, 2012년 이전은 각년도 6월 기준

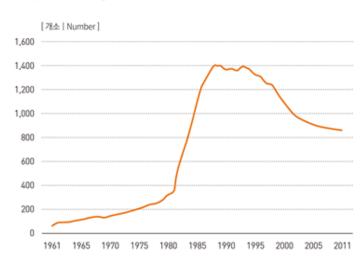
연도별		`07	,08	`09	`10	111	`12	`13	114	`15	`16
영유아	계	530, 922	523, 287	506, 755	502, 633	502, 766	502, 984	489, 543	480,047	472,648	453, 439
인구수 (명)	영아 (0~2)세	261,819	266, 263	260, 370	255, 247	251,951	256, 528	249,819	242,963	232,007	222, 627
	유아 (3~5)세	269, 103	257, 024	246, 385	247, 386	250,815	246, 456	239, 724	237,084	240, 641	230, 812
어린이집	계 :	177,804	180, 178	185, 668	199,651	208, 985	234, 597	238, 581	240,049	235, 089	233, 785
		33,50%	34.40%	36.60%	39.70%	41.60%	46.60%	48.70%	50.00%	49.70%	51.60%
	영아 (0~2)세	76, 699	83, 189	90, 873	104,669	111,495	134, 174	136, 696	139,314	135, 338	131,081
	유아 (3~5)세	101,105	96, 989	94, 795	94, 982	97, 490	100, 423	101,885	100, 735	99, 751	102, 704
유치원		82, 461	79, 274	78, 279	77,075	81,237	87, 997	92, 400	91, 169	91,359	91,016
(3~5)세		15.50%	15.10%	15.40%	15.30%	16.20%	17.50%	18.90%	19.00%	19.30%	20.10%
시설 미이용		270,657	263, 835	242,808	225, 907	212,544	180, 390	158, 562	148, 829	146, 200	128, 638
(0~5)세		51.00%	50.40%	47.90%	44.90%	42.30%	35,90%	32.40%	31.00%	30.90%	28.40%

주제 선정 이유_문제의 원인





그래프 4-1. 유치원 수 1961-2011 Graph 4-1. Kindergartens, 1961-2011



주제 선정 이유_문제의 원인

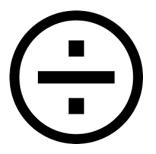






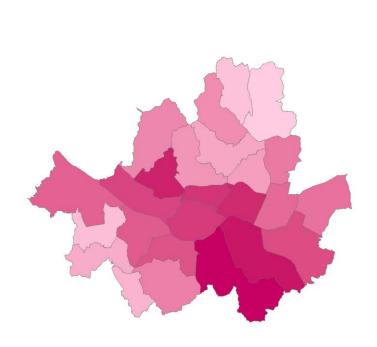
1. 구 선정_영유아인구밀도







1. 구선정_시설당영유아수



자치구별	보육시설 수	영유아 수	시설당 영유아 수
서울시	6,368	349,419	54.87107412
종로구	81	4,167	51.4444444
중구	67	4,308	64.29850746
용산구	125	8,007	64.056
성동구	189	12,587	66.5978836
광진구	217	12,046	55.51152074
동대문구	231	11,409	49.38961039
중랑구	260	13,874	53.36153846
성북구	310	15,238	49.15483871
강북구	182	9,401	51.65384615
도봉구	265	10,841	40.90943396
노원구	494	18,484	37.41700405
은평구	314	16,307	51.93312102
서대문구	159	10,742	67.55974843
마포구	228	14,297	62.70614035
양천구	342	15,723	45.97368421
강서구	438	24,927	56.9109589
구로구	356	17,188	48.28089888
금천구	174	8,318	47.8045977
영등포구	266	15,356	57.72932331
동작구	235	14,296	60.83404255
관악구	281	14,911	53.06405694
서초구	201	16,930	84.22885572
강남구	243	17,866	73.52263374
송파구	421	26,350	62.58907363
강동구	289	15,846	54.83044983

1. 구 선정











2. 변수 설정







Kingdom of Children



영유아수



혼인수

3. 분석 과정_화귀분석

회귀 분석

변수 간의 관계를 파악하여 가중치를 설정

3. 분석 과정 회귀 분석

혼인 수와 출생 수로 영유아 수 예측

model1.lm =lm(kids~marr+birth,finaldata)

summary(model1.lm)

모든 변수가 유의

```
Call:lm(formula = kids ~ marr + birth, data = finaldata)
Residuals:
                 Median 3Q
  Min
         1Q
                                Max
-651.81 -100.74
                 2.19
                        92.07
                               802.74
 Coefficients:
         Estimate Std. Error t value Pr(>|t|)
(Intercept) 128.0479
                     51.6443 2.479
                                      0.0151 *
                      0.4118 -2.120 0.0368 *
           -0.8733
marr
            4.4405
                      0.3127 14.203 <2e-16 ***
birth
Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' 1
Residual standard error: 198 on 88 degrees of freedom
Multiple R-squared: 0.7647, Adjusted R-squared: 0.7593
F-statistic: 143 on 2 and 88 DF, p-value: < 2.2e-16
```

3. 분석 과정 회귀 분석

영유아 수, 혼인 수, 출생 수로 보육시설 수 예측

model2.lm

=lm(kg~kids+marr+birth,finaldata)

summary(model2.lm)

영유아 수 변수만 유의

```
Call:
lm(formula = kg ~ kids + marr + birth, data = finaldata)
Residuals:
                    Median
  Min
           10
                              3Q
                                    Max
                    0.7399 5.6574 16.5194
-17.9469 -5.2480
Coefficients:
     Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.447528 2.086744 -0.694 0.48973
kids
            0.013069 0.004164 3.138 0.00232 **
           -0.009514 0.016494 -0.577 0.56557
marr
birth
            0.042113 0.022161 1.900 0.06070 .
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 7.733 on 87 degrees of freedom
Multiple R-squared: 0.5447, Adjusted R-squared: 0.529
F-statistic: 34.7 on 3 and 87 DF, p-value: 7.614e-15
```

3. 분석 과정 화귀분석

영유아 수로 보육시설 수 예측

 $model3.lm = lm(kg \sim kids, finaldata)$

summary(model3.lm)

Adjusted R-squared가 0.5 근방으로 높지 않은 것으로 볼 때 파악하지 못 한 변수들이 있음

```
Call:
lm(formula = kg ~ kids, data = finaldata)
Residuals:
            10 Median
   Min
                            3Q
                                   Max
-18.8464 -4.9935 0.7629 5.0585 17.8232
Coefficients:
     Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.767686 1.969440 -0.390
kids
           ---Signif. codes: 0 (***, 0.001 (**, 0.01 (*, 0.05 (., 0.1 (), 1
Residual standard error: 7.814 on 89 degrees of freedom
Multiple R-squared: 0.5245, Adjusted R-squared: 0.5192
F-statistic: 98.17 on 1 and 89 DF, p-value: 4.956e-16
```

3. 분석 과정_how

지도학습과 비지도 학습을 혼합하여 분석

5개 구에 대해 영유아 수, 혼인 수, 출생 수를 이용해 정규화 점수를 이용한 군집분석



3. 분석 과정_정규화 점수 계산

Min-Max Normalization

 $100*(x-\min(x))/(\max(x)-\min(x))$

3. 분석 과정_정규화 점수 계산

```
finaldata_3gu_norm
=data_frame(cbind(finaldata_3gu[1],apply(finaldata_3gu[2:4],2,normalize)))
finaldata norm score
=cbind(finaldata_3gu[1],apply(finaldata_3gu_norm[2:4],1,sum))
colnames(finaldata_norm_score) = c("동","score")
finaldata norm score sort
=finaldata_norm_score[order(-finaldata_norm_score$score),]
```

3. 분석 과정_정규화 점수로 군집분석

kmeans_3gu = kmeans(finaldata_norm_score[2],3)

```
> kmeans_3gu = kmeans(finaldata_3gu_norm_score[2],3)
> kmeans_3gu
K-means clustering with 3 clusters of sizes 11, 20, 20
Cluster means:
        score
1 212.52935
2 112.40514
3 61.93638
```

3. 분석 과정_군집 분석 결과 리벨링

상

finaldata_3gu[kmeans_3gu\$cluster == 1,]\$class = "high "

중

finaldata_3gu[kmeans_3gu\$cluster == 2,]\$class = "middle"

하

finaldata_3gu[kmeans_3gu\$cluster == 3,]\$class = "low"

high low middle

3. 분석 과정_의사결정나무 분석

tree1 = ctree(class~kids+marr+birth, data=finaldata_3gu) birth p < 0.001 ≤ 245 > 245 marr p < 0.001 ≤85 >85 4 marr p = 0.01≤116 > 116 Node 7 (n = 10)Node 3 (n = 17)Node 5 (n = 13)Node 6 (n = 11)8.0 8.0 8.0 8.0 0.6 0.6 0.6 0.6 0.4 0.4 0.4 0.4 0.2 0.2 0.2 0.2 0 high low middle high low middle

high low middle

3. 분석 과정_의사결정나무 분석

```
> pl = predict(treel)
> confusionMatrix(finaldata_3gu$class,pl)
Confusion Matrix and Statistics
```

Reference

Prediction high low middle

high 10 0 1 low 0 16 4 middle 0 1 19

Overall Statistics

Accuracy: 0.8824

- > RF = randomForest(class~kids+marr+birth, data=finaldata_3gu, ntree=5000)
 > confusionMatrix(RF\$predicted,finaldata 3gu\$class)
- Gartaria Matrix and Statistics

Confusion Matrix and Statistics

Reference

Prediction high low middle

high 10 0 1 low 0 19 0 middle 1 1 19

Overall Statistics

Accuracy: 0.9412

3. 분석 과정_정규화 점수 계산

```
finaldata_2gu_norm
=data_frame(cbind(finaldata_2gu[1],apply(finaldata_2gu[2:4],2,normalize)))
finaldata norm score
=cbind(finaldata_2gu[1],apply(finaldata_2gu_norm[2:4],1,sum))
colnames(finaldata_norm_score) = c("동", "score")
finaldata norm score sort
=finaldata_norm_score[order(-finaldata_norm_score$score),]
```

```
> finaldata_2gu_norm_score_sort

동 score

71 양재1동 266.64810

92 세곡동 238.91170

82 역삼1동 192.16315

60 잠원동 168.29271

83 역삼2동 130.01521

72 양재2동 127.90529

94 청담동 124.61431

73 내곡동 114.99806

58 서초3동 114.04269

88 일원본동 110.81139
```

3. 분석 과정_정규화 점수로 군집분석

finaldata_norm_score_sort kmeans_3gu = kmeans(finaldata_norm_score[2],3)

```
> kmeans_2gu = kmeans(finaldata_2gu_norm_score[2],3)
> kmeans_2gu
K-means clustering with 3 clusters of sizes 4, 19, 17
Cluster means:
        score
1 216.50391
2 101.36448
3 49.10709
```

3. 분석 과정_군집 분석 결과 라벨링

상

finaldata_2gu[kmeans_2gu\$cluster == 1,]\$class = "high "

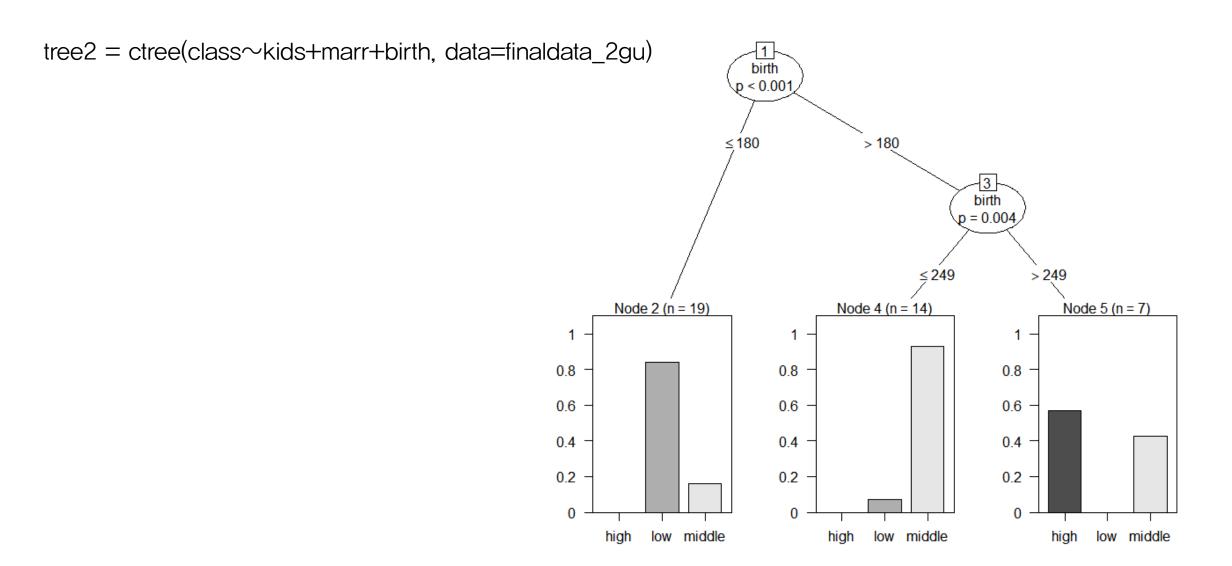
중

Finaldata_2gu[kmeans_2gu\$cluster == 2,]\$class = " middle "

하

finaldata_2gu[kmeans_2gu\$cluster == 3,]\$class = "low"

3. 분석 과정_의사결정나무 분석



3. 분석 과정_의사결정나무 분석

```
> p2 = predict(tree2)
> confusionMatrix(finaldata_2gu$class,p2)
Confusion Matrix and Statistics
```

Reference

Prediction high low middle

high 4 0 0 low 0 16 1 middle 3 3 13

Overall Statistics

Accuracy: 0.825

- > RF = randomForest(class~kids+marr+birth, data=finaldata_2gu, ntree=5000)
- > confusionMatrix(RF\$predicted,finaldata_2gu\$class)
 Confusion Matrix and Statistics

Reference

Prediction high low middle

high 2 0 0 low 0 14 3 middle 2 3 16

Overall Statistics

Accuracy: 0.8

4. 분석 결과

표준화 점수 / 보육시설 수

finaldata_2gu\$ratioscore = finaldata_2gu_norm_score\$score/finaldata_2gu\$kg





3

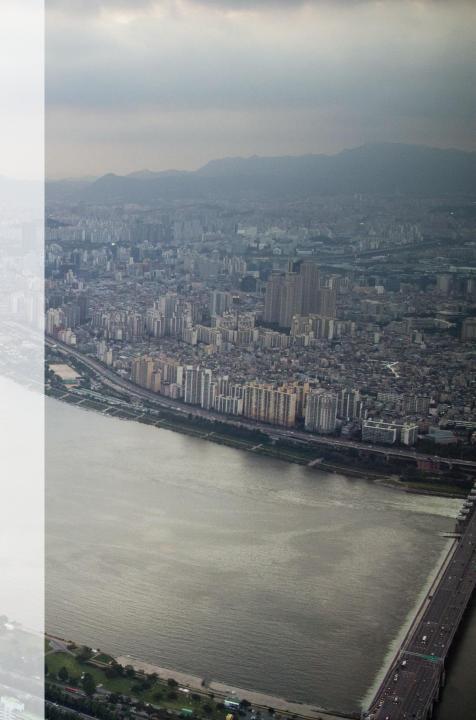
GIS를 사용한 입지 분석

GIS 소개

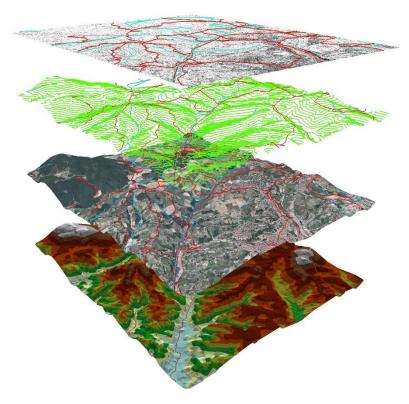
분석 방법 소개

분석 과정 소개

분석 결과



1. GIS 소개_LID기법 정의



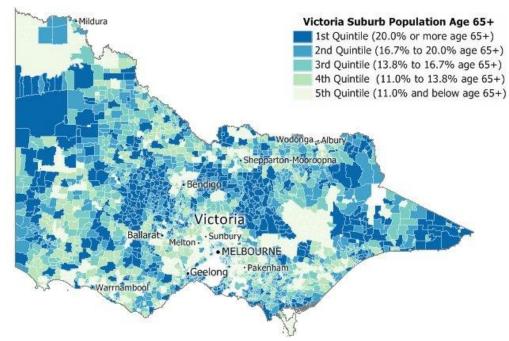
****GIS Layer 모식도 표현

- GIS란

GIS란 'Geographic Information System'의 약자로 인간생활에 필요한 지리정보를 컴퓨터 데이터로 변환하여 효율적으로 활용하기 위한 정보시스템을 말한다.

GIS는 지리적 위치를 갖고 있는 대상에 대한 위치자료와 속성자료를 통합 관리하여 지도, 도표, 및 그림들과 여러 형태의 정보를 제공한다.

1. GIS 소개_LID기법 정의



***실제 GIS분석 예시

- GIS란

GIS란 'Geographic Information System' 의 약자로 인간생활에 필요한 지리정보를 컴퓨터 데이터로 변환하여 효율적으로 활용하기 위한 정보시스템을 말한다.

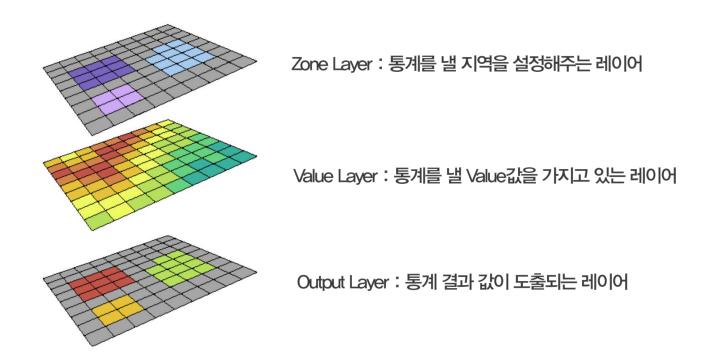
GIS는 지리적 위치를 갖고 있는 대상에 대한 위치자료와 속성자료를 통합 관리하여 지도, 도표, 및 그림들과 여러 형태의 정보를 제공한다.

지리정보를 이용한 빅데이터 분석 시스템

그렇다면 아무 드로 가는 분석할 것인가?

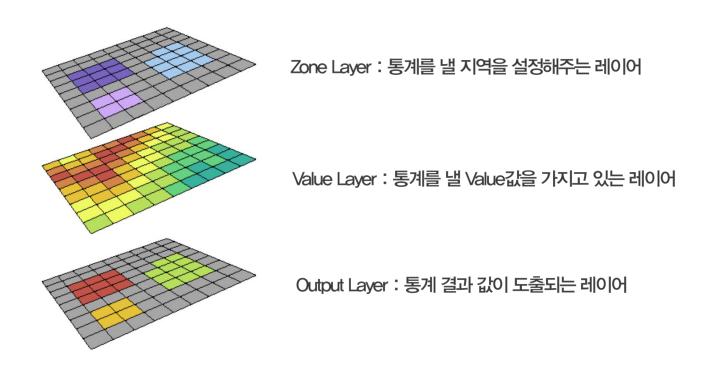
2. 분석 방법 소개_Zonal Statistics

*Zonal Statistics: 다른 데이터 셋의 영역 내에서 래스터 값에 대한 통계를 계산. 입력 영역 데이터의 각 영역별로 하나의 결과 값이 계산된다.

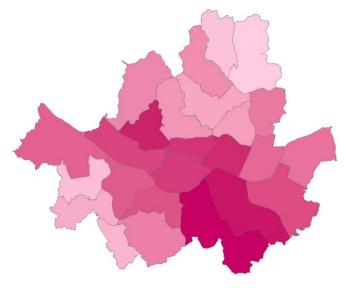


2, 분석 방법 소개_Zonal Statistics

*Zonal Statistics: 다른 데이터 셋의 영역 내에서 래스터 값에 대한 통계를 계산. 입력 영역 데이터의 각 영역별로 하나의 결과 값이 계산된다.



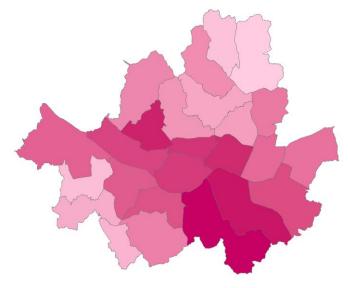
다양한통계값을이용한여러개의 layer를 만든후에 "Zonal을 통해 데이터를 합산하여 최적의 위치를 결정한다."



***서울시 구별 어린이집 밀도

1. 서울시 구별 어린이집 밀도

서울시에서 인구대비 보육시설이 가장 적은 하위 2개의 구를 골라 분석의 범위를 좁힌다.



***서울시 구별 보육시설 당 이동 수

1. 서울시 구별 어린이집 밀도

서울시에서 인구대비 보육시설이 가장 적은 하위 2개구를 골라 분석의 범위를 좁힌다.



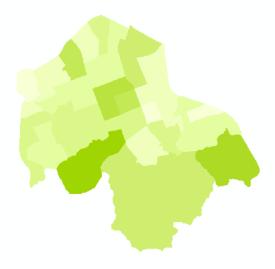
***하위 2개 자치구 추출

서울시	27,4355								
종로구	25,7222	동대문구	24.6948	노원구	18,7085	강서구	28,4555	관악구	26,532
중구	32,1493	중랑구	26,6808	은평구	25,9666	구로구	24.1404	서초구	42.1144
용산구	32,028	성북구	24.5774	서대문구	33.7799	금천구	23,9023	강남구	36,7613
성동구	33,2989	강북구	25.8269	마포구	31,3531	영등포구	28.8647	송파구	31,2945
광진구	27.7558	도봉구	20.4547	양천구	22,9868	동작구	30.417	강동구	27.4152

〈서울시 자치구별 보육시설 당 이동 수〉

서초구와 강남구가 하나의 보육원에 가장 많은 아동 수를 수용하는 것으로 나타남

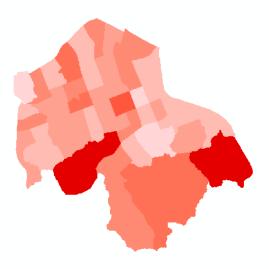
3. 분석 과정 소개_레스터화를 통한 분류



〈서초구/강남구 동 별 출생 수〉

Value

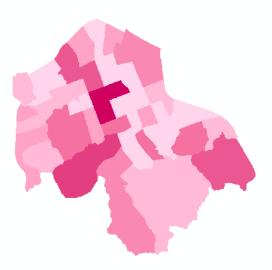
- <u>62 105.1111111</u>
- **105.11111112 148.2222222**
- 148.2222223 191.3333333
- <u>191.3333334 234.4444444</u>
- 234.4444445 277.5555556
- 277.5555557 320.6666667
- 320.6666668 363.7777778
- 363.7777779 406.8888889
- 406.888889 450



〈서초구/강남구 동 별 영유아(0~4세) 수〉

Value

- 236 474.3333333
- **474.3333334 712.6666667**
- 712.6666668 951
- **951.0000001 1,189.333333**
- **1,189.333334 1,427.666667**
- 1,427.666668 1,666
- **1**,666.000001 1,904.333333
- **1,904.333334 2,142.666667**
- **2,142.666668 2,381**

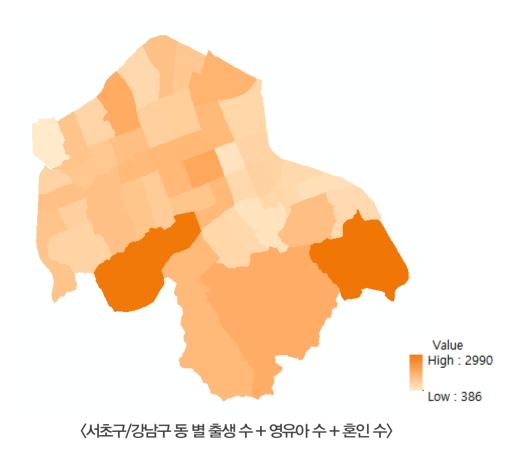


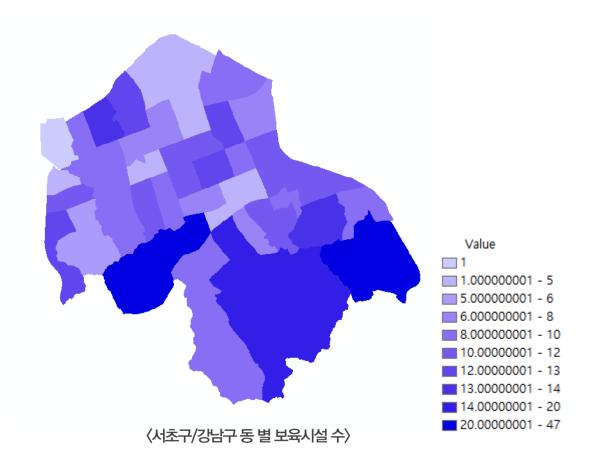
〈서초구/강남구 동 별 혼인 수〉

Value

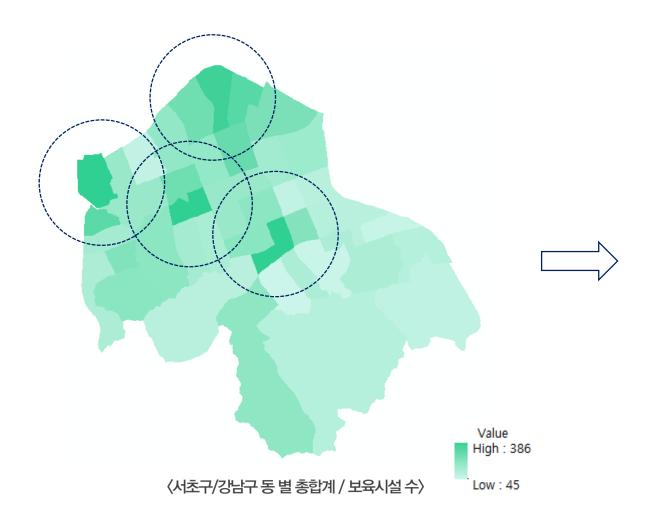
- 28 67.333333333
- Fig. 67.33333334 106.6666667
- **106.6666668 146**
- **146.0000001 185.3333333**
- **185.3333334 224.6666667**
- **224.6666668 264**
- **264.0000001 303.3333333**
- 303.3333334 342.6666667
- **342.6666668 382**

3. 분석 과정 소개_레스터의 합계



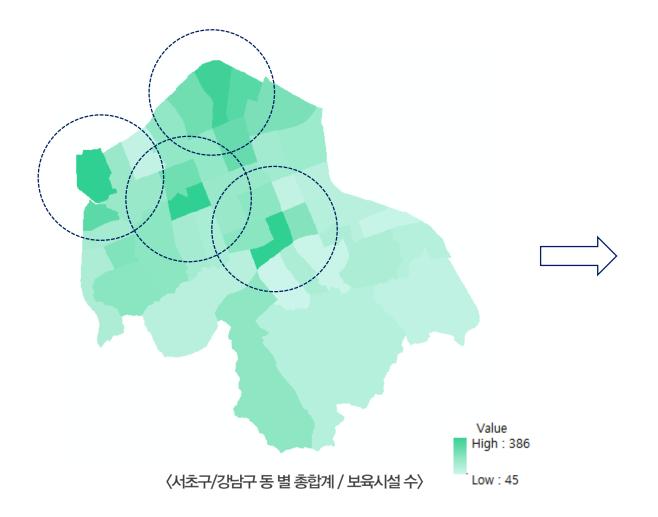


3. 분석 과정 소개_레스터의 합계



원 안에 있는 부분이 보육시설 당 아동 수가 많은 지역 따라서 이 동그라미 친 부분에 보육시설을 추가로 설치할 필요성이 제기된다.

3. 분석 과정 소개_지역 추출

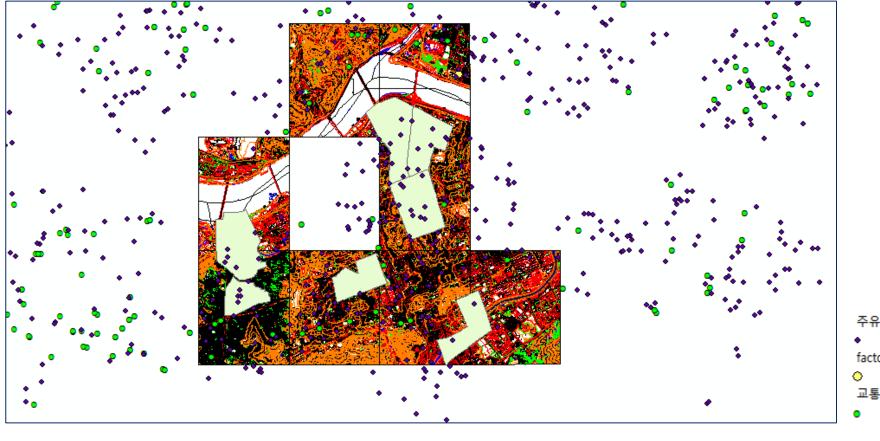


> finaldata_2gu_ratioscore_sort

원 안에 있는 부분이 보육시설 당 이동 수가 많은 지역

따라서 이 동그라미 친 부분에 보육시설을 추가로 설치할 필요성이 제기된다.

> 반포본동 논현2동 압구정동 도곡2동

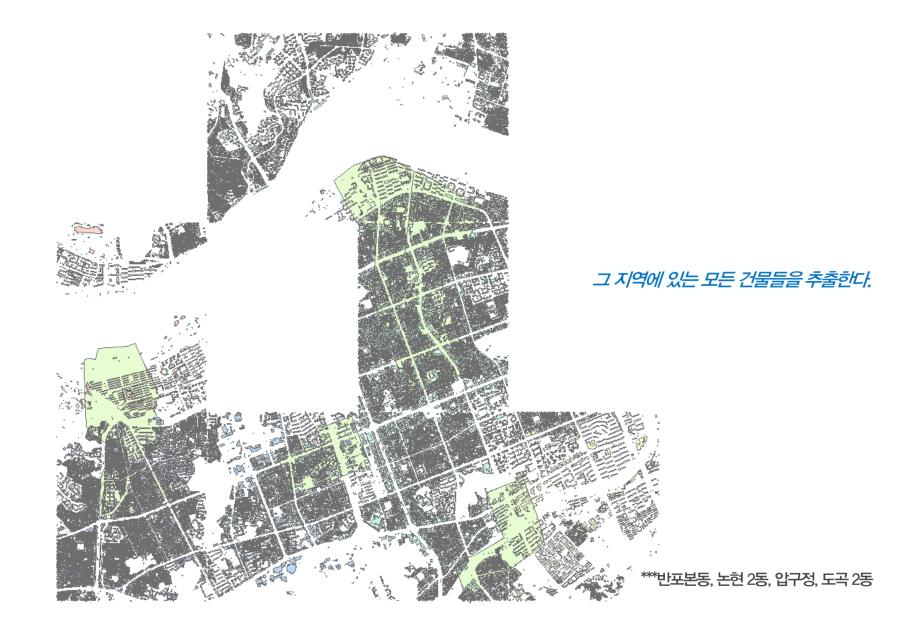


주유소, 공장, 사고다발지역, 수치지도를 추출한 동 위에 겹친다

주유소

factory

교통사고다발지역





그 지역에 있는 모든 건물들을 추출한다.

***반포본동, 논현 2동, 압구정, 도곡 2동

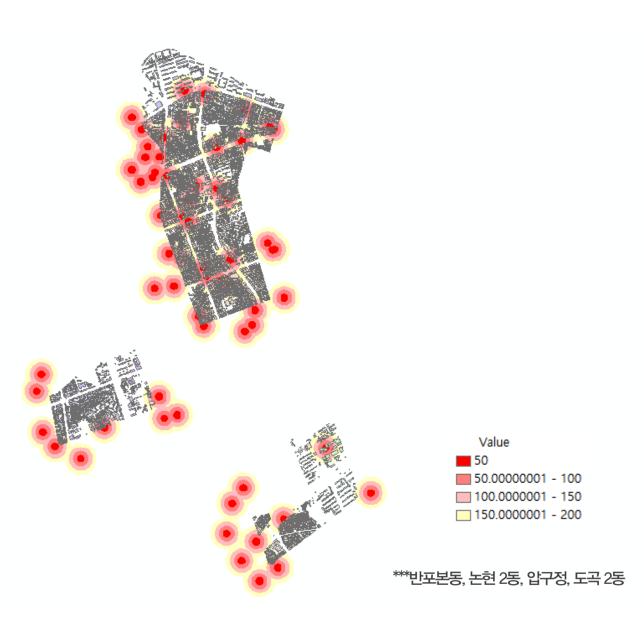
3. 분석 과정 소개_위험 지역 범위 설정

"공동주택, 어린이놀이터, 의료시설, 유치원, 어린이집 및 경로 당은 정부 법령 상 주유소나 공장에서 수평거리 50m 내에는 설 치할 수 있다."

> "그러나 최근 어린이 보호의 필요성 증대와 이미 확보되어 있는 어린이집 밀도를 고려하여 수평거리 50m를 200m로 늘리는 것이 좋다."

> > 육아정책연구소(양미선, 임지희)





4. 분석 결과_결과 값의 입력





R

```
normalize \langle - function(x) \rangle
return (100*(x-min(x))/(max(x)-min(x)))
```

 $MIN \rightarrow 0 / MAX \rightarrow 100$

GIS

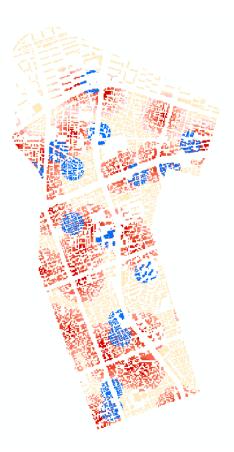
Jenks Natural Breaks Classification

데이터 값의 배열을 자연스러운 등급으로 최적화 등급 내의 분산 최소화, 등급 간의 분산 최대화

최종결과

R

GIS



Thank you () - !



