

각세별 산모연령 출산순위별 출생아수에 대한 탐색적 시각화분석[†]

김종태¹

¹대구대학교 전산통계학과

접수 2017년 12월 20일, 수정 2018년 1월 9일, 게재확정 2018년 1월 9일

요 약

본 연구는 통계청 국가통계포털의 인구동향조사 ‘각 세별 모의연령 및 출산순위별 출생아수’ 데이터를 가지고 다양한 시각화 도구들을 이용한 시각화 분석으로 출생아수에 대한 통계적 분포를 분석하는 것이 목적이다. 본 연구에서는 가장 쉽게 데이터를 그래픽화 하여 사용할 수 있는 마이크로소프트사의 엑셀과 무료이고 오픈 소프트웨어인 R을 사용하여 탐색적인 통계분석 방법을 구현하였다. 다양한 삼차원 그래프들과 그래프의 이동각도들의 변화에 따른 시각화 분석들에 비교와 이차원 그래프들에 따른 분석, 파이그래프나 막대그래프 등의 여러 종류의 시각화도구들을 가지고 연도별 연령별 변화에 따른 출생 분포의 변화들을 비교를 하였다. 다양한 탐색적 시각화 분석을 통하여 출산순위별이나 모의 연령별에 따른 출생 데이터의 분포와 출생아수를 예측할 수 있다.

주요용어: 국가통계포털, 시각화분석, 출산 순위별 출생.

1. 서론

최근 낮은 출산율은 국가적인 재앙을 예견할 만큼 위험한 수준에 이르렀다. 이에 정부는 출산 장려를 위하여 수조원의 예산을 투입하지만 출산율은 오히려 더 떨어지고 있는 현실에 직면하고 있다. 이에 본 연구에서는 통계청 국가통계포털 (Korean statistical information service; KOSIS)의 인구동향조사의 ‘성/모의 연령 (각세)/출산순위별 출생’ (live births by age group of mother, sex and birth order period annual 1981 ~ 2016)에 대한 데이터를 시각화 분석을 위한 다양한 그래프들을 가지고 ‘성/모의 연령 (각세)/출산순위별 출생’ 데이터들을 분석하는데 목적이 있다. 본 연구에서 제시한 시각화 그래프들은 통계를 이해하고자 하는 학부학생들의 탐색적 데이터분석 분야의 시각화 교육에 활용할 수 있을 것으로 기대한다.

데이터 과학 분야에 있어서 시각화 방법론은 분석대상인 데이터를 관찰하는 방법이나 표현하는 방법에 따라서, 데이터가 가지고 있는 본래의 모습들을 데이터 스스로가 스스로 표현하게 하는 기법 중 하나이다. 시각화 분석은 최근 빅데이터 분석의 발달로 매우 다양한 그래픽 분석 방법들을 제시하거나 혹은 새롭게 개발되고 창조되고 진화된 시각화 도구들이 만들어지고 있다.

통계표명 ‘성/모의 연령 (각세)/출산순위별 출생’ 데이터는 통계청의 인구동향조사 의 2017년 11월 27일 통계로서 수록기간은 1981년부터 2016년까지이며 통계표 ID는 DT_1B80A01이다 (KOSIS,

[†] 이 논문은 2015년 대구대학교 학술연구비 지원에 의한 연구임.

¹ (38453) 경북 경산시 진량읍 대구대로 201, 대구대학교 전산통계학과, 교수.

E-mail: jtkim@daegu.ac.kr

2017). 통계표는 산모 연령은 15세 미만에서부터 50세 이상까지와 연령미상을 포함하고 있다. 출생 순위는 1아부터 8아 이상까지 포함되어 있고, 남자, 여자, 총계로 분류해서 조사했다.

시각화분석에 사용할 수 있는 소프트웨어나 도구들은 매우 다양하게 많다. 그 중에 시각화 방법론은 가장 쉽게 사용할 수 있는 마크로소프트사의 엑셀과 R을 이용하여 구현하였다. 마이크로소프트웨어사의 엑셀은 그 기능면에서 사용하기 매우 쉽고 뛰어난 그래픽 기능을 가지고 있다. 또한 대부분의 PC에서 설치되어 있으므로 쉽게 사용되며 분석이 용이하여 산업현장에서 매우 널리 사용되고 통계분석에서도 매우 유용하게 사용된다. 무료이자 오픈소스인 R은 그래픽 기능면에서 매우 우수한 시각화 분석도구이다.

Ahn 등 (2017)은 고차원 자료에서 영향점을 평가하기 위한 그래픽 방법들을 제안하고 고차원 실제 자료에서 그래픽 방법들을 적용하였다. Lee (2017)은 통계청과 한국은행의 경제통계 시각화의 특성을 정리하고 언론 등의 다양한 경제통계 시각화 결과를 바탕으로 향후방안을 제시하였다. Na와 Hwang (2017)은 R의 'shiny' 패키지를 이용하여 공간 자료와 관계된 분석으로 등고선 등의 형태로 표현되는 지형자료를 분석하여 지질 이상대를 추출하고, 기상, 환경, 소셜미디어 정보를 이용하여 전국의 16개 시, 도별 주요 질환을 예측하는 모형을 구축하였다. Yoo (2017)은 제조 빅데이터 시스템에서 데이터 수집, 분석 및 예측 결과를 효과적으로 보여 주기 위해 사용가능한 시각화 기법을 제시하였다.

다차원 데이터의 분석들은 다양한 그래픽에서의 분석방법을 이용하여 시간시각화, 관계시각화, 비교 시각화, 분포시각화와 공간시각화 기법으로 구분되어 데이터 분석 및 예측 결과를 보다 효과적으로 시각화할 수 있는 분석 방법을 제공한다. 본 연구는 2절에서는 이차원 그래픽에 의한 시각화 분석과 3절에서는 삼차원 그래픽에 의한 시각화 분석 4절에서는 분석과 결론으로 구성되어 있다.

2. 각세별 산모연령별 출생아수의 이차원 시각화 분석

통계청의 인구동향조사 (KOSIS, 2017)의 ‘성/모의 연령 (각세)/출산순위별 출생’ 데이터는 산모연령 15세 미만에서부터 50세 이상까지와 연령미상을 포함하고 있다. 출생순위는 1아부터 8아 이상까지 포함되어 있고, 남자, 여자 그리고 총계로 분류해서 1981년부터 2016년까지 아래의 Figure 2.1과 같이 조사되었다.

Life Events by Age Group of Father, Sex and Order of Birth		1957		1958		1959		1960		1961		1962		1963		1964		1965		1966		1967		1968		1969		1970		1971		1972		1973		1974		1975		1976		1977		1978		1979		1980		1981		1982		1983		1984		1985		1986		1987		1988		1989		1990		1991		1992		1993		1994		1995		1996		1997		1998		1999		2000		2001		2002		2003		2004		2005		2006		2007		2008		2009		2010		2011		2012		2013		2014		2015		2016		2017		2018		2019		2020		2021		2022		2023		2024		2025		2026		2027		2028		2029		2030		2031		2032		2033		2034		2035		2036		2037		2038		2039		2040		2041		2042		2043		2044		2045		2046		2047		2048		2049		2050		2051		2052		2053		2054		2055		2056		2057		2058		2059		2060		2061		2062		2063		2064		2065		2066		2067		2068		2069		2070		2071		2072		2073		2074		2075		2076		2077		2078		2079		2080		2081		2082		2083		2084		2085		2086		2087		2088		2089		2090		2091		2092		2093		2094		2095		2096		2097		2098		2099		2100		2101		2102		2103		2104		2105		2106		2107		2108		2109		2110		2111		2112		2113		2114		2115		2116		2117		2118		2119		2120		2121		2122		2123		2124		2125		2126		2127		2128		2129		2130		2131		2132		2133		2134		2135		2136		2137		2138		2139		2140		2141		2142		2143		2144		2145		2146		2147		2148		2149		2150		2151		2152		2153		2154		2155		2156		2157		2158		2159		2160		2161		2162		2163		2164		2165		2166		2167		2168		2169		2170		2171		2172		2173		2174		2175		2176		2177		2178		2179		2180		2181		2182		2183		2184		2185		2186		2187		2188		2189		2190		2191		2192		2193		2194		2195		2196		2197		2198		2199		2200		2201		2202		2203		2204		2205		2206		2207		2208		2209		2210		2211		2212		2213		2214		2215		2216		2217		2218		2219		2220		2221		2222		2223		2224		2225		2226		2227		2228		2229		2230		2231		2232		2233		2234		2235		2236		2237		2238		2239		2240		2241		2242		2243		2244		2245		2246		2247		2248		2249		2250		2251		2252		2253		2254		2255		2256		2257		2258		2259		2260		2261		2262		2263		2264		2265		2266		2267		2268		2269		2270		2271		2272		2273		2274		2275		2276		2277		2278		2279		2280		2281		2282		2283		2284		2285		2286		2287		2288		2289		2290		2291		2292		2293		2294		2295		2296		2297		2298		2299		2300		2301		2302		2303		2304		2305		2306		2307		2308		2309		2310		2311		2312		2313		2314		2315		2316		2317		2318		2319		2320		2321		2322		2323		2324		2325		2326		2327		2328		2329		2330		2331		2332		2333		2334		2335		2336		2337		2338		2339		2340		2341		2342		2343		2344		2345		2346		2347		2348		2349		2350		2351		2352		2353		2354		2355		2356		2357		2358		2359		2360		2361		2362		2363		2364		2365		2366		2367		2368		2369		2370		2371		2372		2373		2374		2375		2376		2377		2378		2379		2380		2381		2382		2383		2384		2385		2386		2387		2388		2389		2390		2391		2392		2393		2394		2395		2396		2397		2398		2399		2400		2401		2402		2403		2404		2405		2406		2407		2408		2409		2410		2411		2412		2413		2414		2415		2416		2417		2418		2419		2420		2421		2422		2423		2424		2425		2426		2427		2428		2429		2430		2431		2432		2433		2434		2435		2436		2437		2438		2439		2440		2441		2442		2443		2444		2445		2446		2447		2448		2449		2450		2451		2452		2453		2454		2455		2456		2457		2458		2459		2460		2461		2462		2463		2464		2465		2466		2467		2468		2469		2470		2471		2472		2473		2474		2475		2476		2477		2478		2479		2480		2481		2482		2483		2484		2485		2486		2487		2488		2489		2490		2491		2492		2493		2494		2495		2496		2497		2498		2499		2500		2501		2502		2503		2504		2505		2506		2507		2508		2509		2510		2511		2512		2513		2514		2515		2516		2517		2518		2519		2520		2521		2522		2523		2524		2525		2526		2527		2528		2529		2530		2531		2532		2533		2534		2535		2536		2537		2538		2539		2540		2541		2542		2543		2544		2545		2546		2547		2548		2549		2550		2551		2552		2553		2554		2555		2556		2557		2558		2559		2560		2561		2562		2563		2564		2565		2566		2567		2568		2569		2570		2571		2572		2573		2574		2575		2576		2577		2578		2579		2580		2581		2582		2583		2584		2585		2586		2587		2588		2589		2590		2591		2592		2593		2594		2595		2596		2597		2598		2599		2600		2601		2602		2603		2604		2605		2606		2607		2608		2609		2610		2611		2612		2613		2614		2615		2616		2617		2618		2619		2620		2621		2622		2623		2624		2625		2626		2627		2628		2629		2630		2631		2632		2633		2634		2635		2636		2637		2638		2639		2640		2641		2642		2643		2644		2645		2646		2647		2648		2649		2650		2651		2652		2653		2654		2655		2656		2657		2658		2659		2660		2661		2662		2663		2664		2665		2666		2667		2668		2669		2670		2671		2672		2673		2674		2675		2676		2677		2678		2679		2680		2681		2682		2683		2684		2685		2686		2687		2688		2689		2690		2691		2692		2693		2694		2695		2696		2697		2698		2699		2700		2701		2702		2703		2704		2705		2706		2707		2708		2709		2710		2711		2712		2713		2714		2715		2716		2717		2718		2719		2720		2721		2722		2723		2724		2725		2726		2727		2728		2729		2730		2731		2732		2733		2734		2735		2736		2737		2738		2739		2740		2741		2742		2743		2744		2745		2746		2747		2748		2749		2750		2751		2752		2753		2754		2755		2756		2757		2758		2759		2760		2761		2762		2763		2764		2765		2766		2767		2768		2769		2770		2771		2772		2773		2774		2775		2776		2777		2778		2779		2780		2781		2782		2783		2784		2785		2786		2787		2788		2789		2790		2791		2792		2793		2794		2795		2796		2797		2798		2799		2800		2801		2802		2803		2804		2805		2806		2807		2808		2809		2810		2811		2812		2813		2814		2815		2816		2817		2818		2819		2820		2821		2822		2823		2824		2825		2826		2827		2828		2829		2830		2831		2832		2833		2834		2835		2836		2837		2838		2839		2840		2841		2842		2843		2844		2845		2846		2847		2848		2849		2850		2851		2852		2853		2854		2855		2856		2857		2858		2859		2860		2861		2862		2863		2864		2865		2866		2867		2868		2869		2870		2871		2872		2873		2874		2875		2876		2877		2878		2879		2880		2881		2882		2883		2884		2885		2886		2887		2888		2889		2890		2891		2892		2893		2894		2895		2896		2897		2898		2899		2900		2901		2902		2903		2904		2905		2906		2907		2908		2909		2910		2911		2912		2913		2914		2915		2916		2917		2918		2919		2920		2921		2922		2923		2924		2925		2926		2927		2928		2929		2930		2931		2932		2933		2934		2935		2936		2937		2938		2939		2940		2941		2942		2943		2944		2945		2946		2947		2948		2949		2950		2951		2952		2953		2954		2955		2956		2957		2958		2959		2960		2961		2962		2963		2964		2965		2966		2967		2968		2969		2970		2971		2972		2973		2974		2975		2976		2977		2978		2979		2980		2981		2982		2983		2984		2985		2986	
--	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--

Figure 2.1 Live births by age group of mother, sex and birth order period annual 1981 ~ 2016

시간의 시각화 관점에서 Figure 2.2는 1981년부터 2016년까지의 출생아 총계에 대한 그래프이다. 연도별 총 출생아수는 1981년 867,409명에서 35년 후인 2016년에는 406,243명으로 1981년 대비 53.2%가 감소하였다. Figure 2.2에서 1982년-1984년 사이와 2000년 - 2002년 사이에 출생아 수가 급감함을 보인다. 그러나 연도별 총 출생아수에 대한 그래프로 출생의 분포를 설명하기엔 부족한 점이 많다.

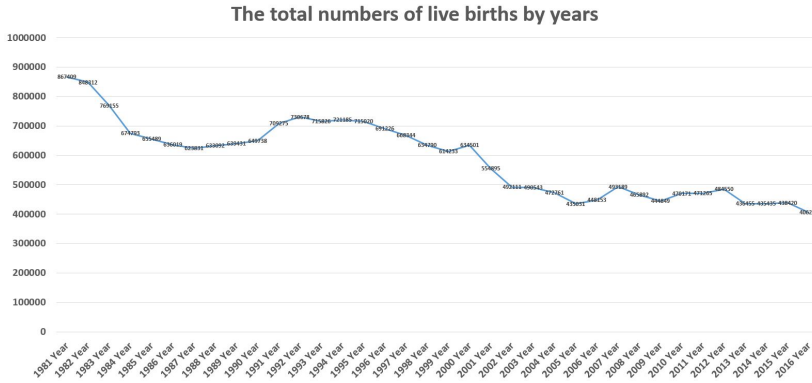


Figure 2.2 The total numbers of live births by years

아래의 Figure 2.3은 1981년부터 2016년까지 각 연도마다 15세 미만에서 50세 이상의 산모연령들의 연령순서로 태어난 출생아들의 크기들을 기준으로 만든 그래프이다. 산모연령들의 구간이 클수록 출생아수가 상대적으로 많게 표현하였다. Figure 2.3은 Figure 2.2의 그래프를 산모연령대의 출생아수 크기에 대한 정보를 첨가한 그래프이다. 1981년 산모연령 25세 전후의 출생아수는 2016년에 가까워짐에 따라 급격히 줄어지고 있음을 보인다. 최근 들어서 25세 산모들의 출생아수가 매우 적다는 것을 의미한다.

Figure 2.4는 시간의 시각화 관점에서 1981년부터 2016년까지 가장 많은 출생아수를 가진 산모들의 연령 순서대로 그린 그래프이다. 1981년 25세 산모들의 출생아수는 최고로 많았으나 2016년에는 이르러서는 급격히 떨어졌고, 1981년 33세의 산모들의 출생아수는 매우 적었으나 2016년에는 가장 높은 출생아수를 기록하고 있다. 이는 산모연령 25세 전후의 출생아수는 1981년 무렵에는 가장 많았지만 2016년에는 매우 저조한 것으로 나타난다. 반면에 32세 전후의 산모연령의 출생아수는 1981년 무렵에는 매우 적었으나 2016년 무렵에는 매우 많은 출생아수를 나타내고 있다. 출생아수가 가장 많은 산모들의 연령대가 25세에서 32세로 점점 이동을 하고 있는 분포의 모양을 보이고 있다.

Figure 2.5는 가장 높은 출생아수의 비율 순으로 산모연령들에 대한 데이터를 연도별로 나타낸 것이다. 1981년 25세 산모들의 출생아수 비율은 11.11%이고 1986년에는 14.14%으로 최고점을 찍었으나 그 후 급격히 떨어져서 2016년에는 2.02%의 비중을 나타낸다. 1981년에는 32세의 산모에게서 난 출생아수의 비율이 1.75%으로 매우 낮았으나 2016년에는 10.34%으로 가장 높은 출생아 비율을 나타내고 있다. 향후 산모연령 32세 이상에서 가장 높은 출생아수 비율을 나타낼 것으로 관측된다.

아래의 Figure 2.6은 위의 Figure 2.5의 꺾은선 비율 그래프보다 보다 명확하게 산모연령에 따른 출생아수의 비율들의 변화를 볼 수 있다. 1981년에는 산모연령 25세가 11%로 가장 높은 출산율을 나타내고 있으나 2016년을 기점으로 산모연령 32세가 전체 10%의 가장 높은 출산율을 나타내고 있다.

Figure 2.7은 15세 미만에서부터 50세 이상의 산모들의 연령대에서 태어난 출생아수에 대한 상자그림들을 1981년부터 2016년까지 나타낸 그래프이다. 이상치들을 포함한 분포모양은 위의 Figure 2.2와 유사한 모양을 나타내고 있다. 그러나 이상치를 제외하고 사분위수의 크기들로 산포의 정도를 살펴보면

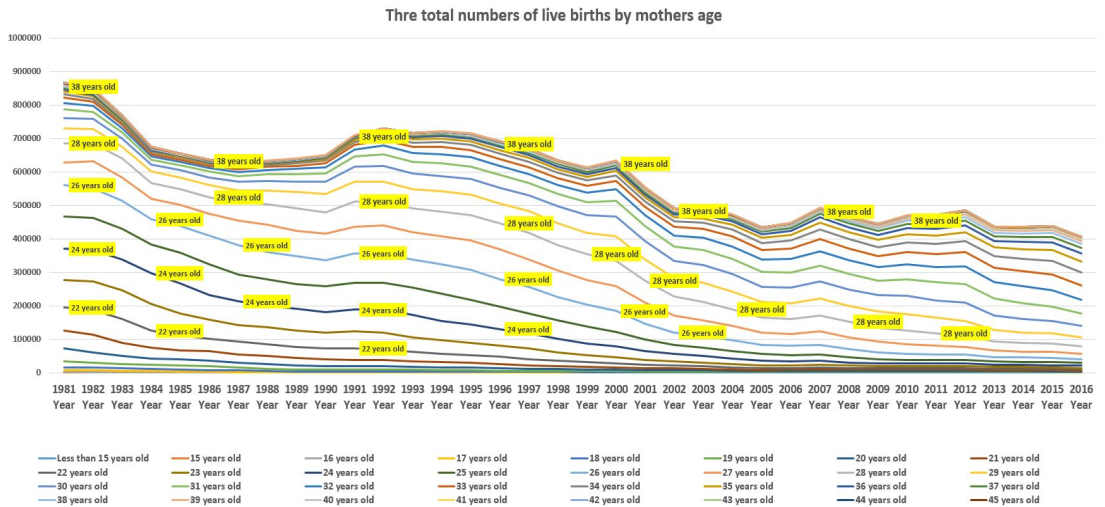


Figure 2.3 The total numbers of live births by mothers age

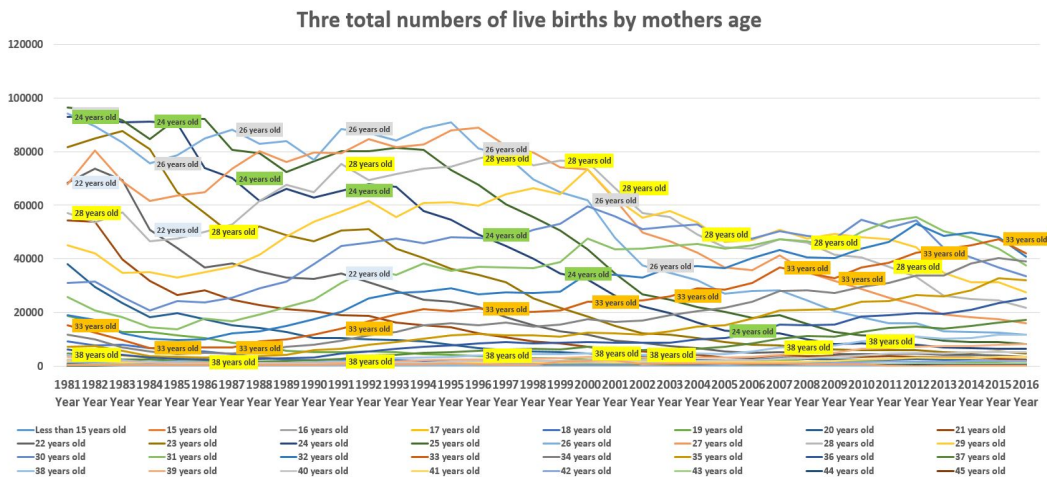


Figure 2.4 The total numbers of live births by mothers age

최근 들어 산모의 연령에 따른 산포는 점점 작아지는 것으로 나타나고 있다.

아래의 Figure 2.8은 시간과 출생순위의 시각화 관점에서 1아부터 8아 이상의 출생순위를 그래프로 표현 하였다. 1981년에 1아의 출생아수가 355,314명이 2016년에 212,932명으로 줄었다. 이어서 2아와 3아, 4아의 출생아수들 역시 2016년에 가까워짐에 따라서 전반적으로 감소하는 현상을 나타내고 있다.

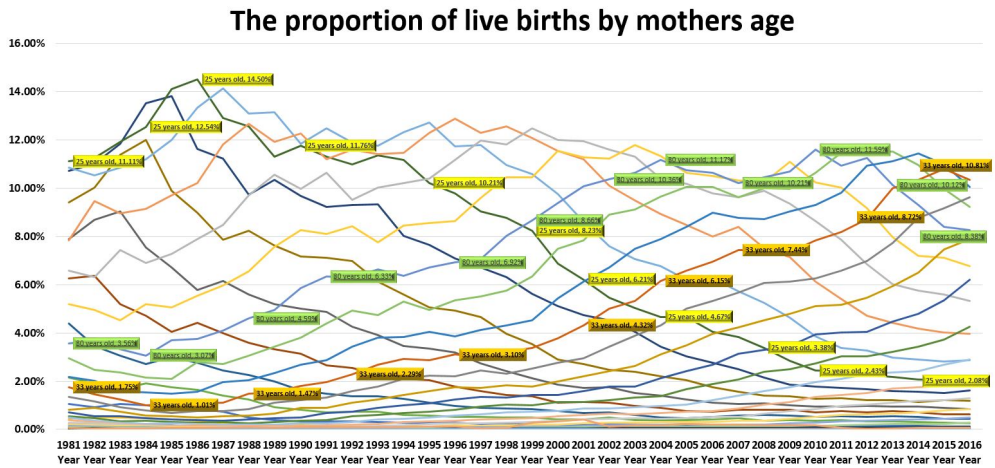


Figure 2.5 The proportion of live birth mothers age

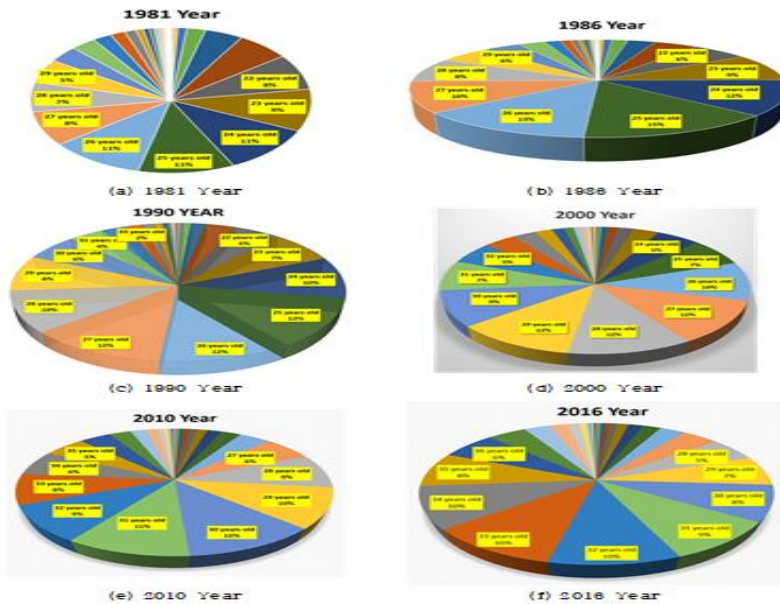


Figure 2.6 The pie-charts for the proportion of live birth mothers age

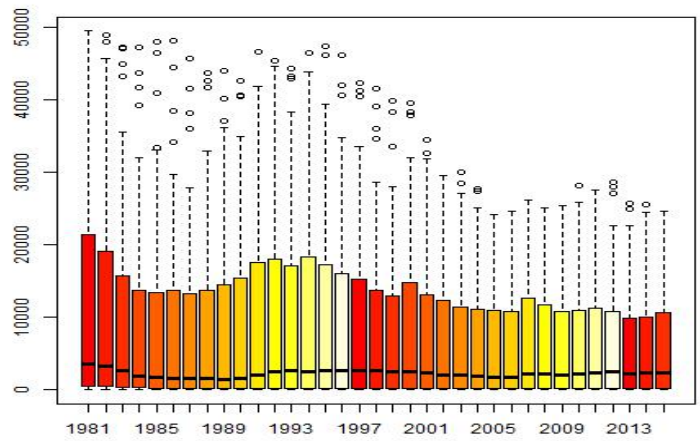


Figure 2.7 Boxplot for the numbers of live births by mother ages

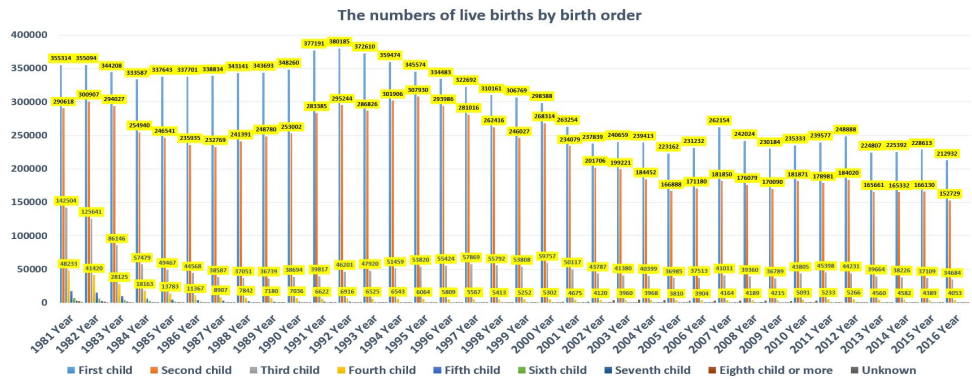


Figure 2.8 The numbers of live births by birth order

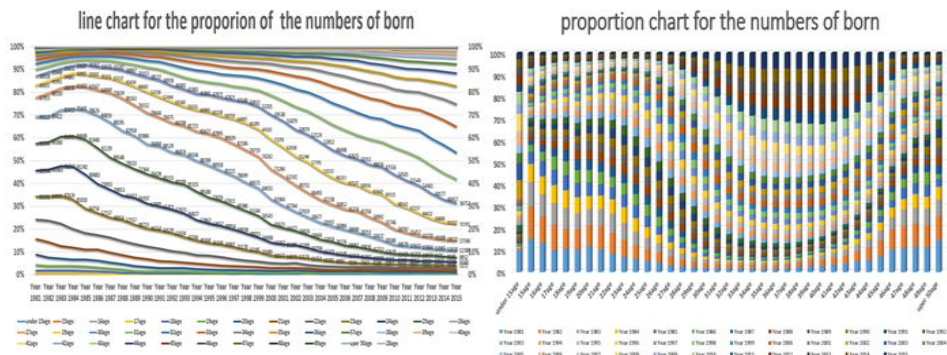


Figure 2.9 proportion charts for the number of live births

Figure 2.9는 1981년부터 2016년까지 연도별로 산모연령별에 따른 출생아수 비율들을 그래프화한 방법들을 나타낸 것이다. 전체 출생아수 대비 산모 연령별 출생아수의 비율 관점에서 볼 수 있는 시각화 그래프들을 표현하였다.

3. 각세별 산모연령별 출생아수의 분포의 시각화 분석

3절에서는 산모연령과 연도에 따른 출생아수를 삼차원 그래프를 이용한 시각화 분석을 하였다. 그래프의 표현방법에 따라서 연도와 산모연령별 출생아수의 분포들이 시각적으로 어떻게 달라지는가를 관찰하고 분석한다. 이러한 분석은 미래 총 인구수를 예측하는 출생아수의 분포의 변화를 탐색적으로 분석할 수 있다.

아래 Figure 3.1은 산모연령에 따른 연도별 출생아수를 막대그래프로 표시한 것이다. 가장 앞쪽의 막대그래프가 2016년의 것으로 32세 산모에 의한 출생아 수가 가장 높게 나오지만, 가장 뒤쪽에 있는 1981년의 그래프는 25세 산모들의 출생아수가 높은 것으로 나타난다. 36년 동안 분포의 중심축이 25세에서 32세로 변해오는 과정을 막대 그래프가 보여주고 있다.

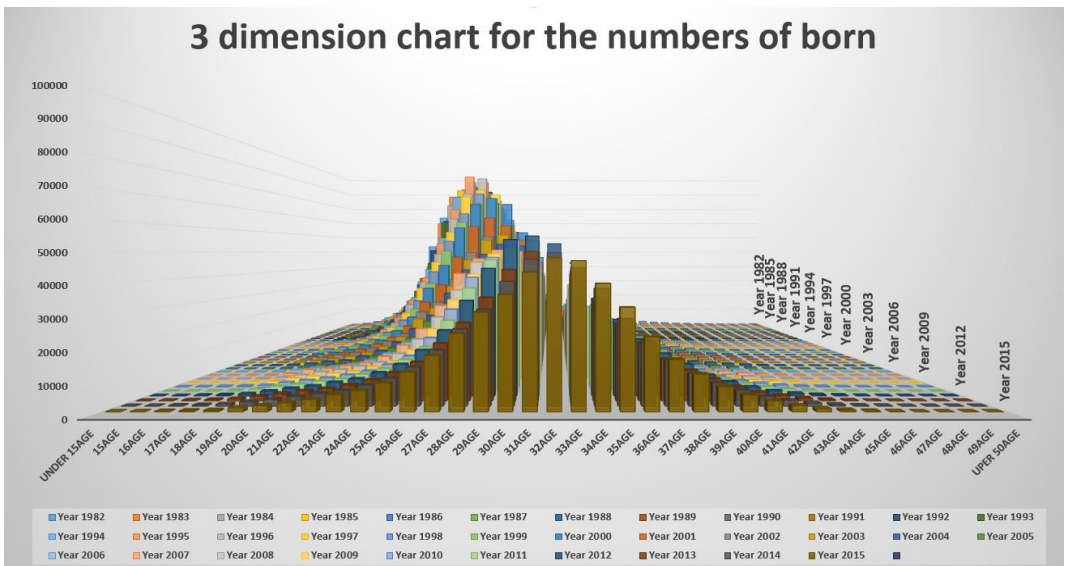


Figure 3.1 3-dimensional bar chart for the number of live births

Figure 3.2는 위의 Figure 3.1을 x -축을 우측으로 15도정도 비튼 삼차원 막대그래프이다. 위의 Figure 3.1 보다도 훨씬 더 명확하게 왼쪽 측면에서부터 오른쪽 방향으로 그래프를 관찰함으로써 출생아수에 대한 분포의 변화가 보다 선명하게, 시각적 효과를 나타내고 있다. 가장 뒤쪽의 1981년 출생아수의 분포가 2016년의 막대의 높이가 높고 퍼짐정도가 큰 것으로 보아 분포 보다 훨씬 높고 큰 규모임을 알 수 있다. 중심위치가 2016년의 32세보다 왼쪽인 25세에 위치하고 있음을 알 수 있다. 이 그래프로 우리는 2017년 출생아수의 분포의 모양과 중심위치는 막대의 높이가 낮아지면서 예측할 수 있을 것이다.

Figure 3.3는 산모연령에 따른 연도별 출생아수의 분포를 방사선 그래프로 표시한 것이다. 이 방사선 그래프의 해석은 0을 중심으로 각 산모들 연령좌표 방향으로 연결한 방사선 형태로 관찰하는 그래프이

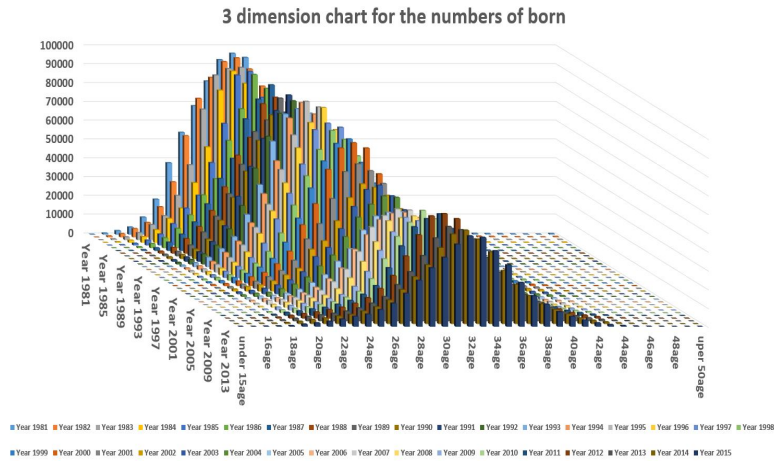


Figure 3.2 3-dimensional bar chart for the number of live births by moving x

다. 가장 왼쪽의 가장 작은 원모양의 그래프가 2016년의 그래프로서 그 산포의 형태가 둥근 원에 가깝고 크기가 작은 원임을 알 수 있다. 오른쪽으로 갈수록 원들은 점점 커지고 그 모양도 타원에 가깝게 커지는 것을 볼 수 있다. 이는 분포가 좌우대칭에 가까울수록 원모양이 되고, 분포가 한쪽으로 많이 치우쳐질수록 점점 타원형으로 변해 감을 알 수 있다.

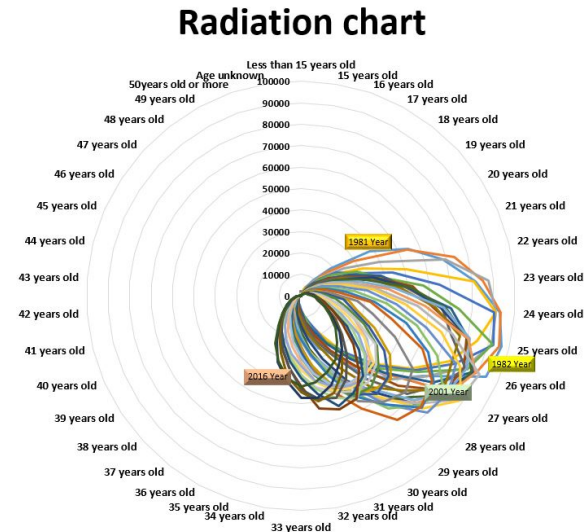


Figure 3.3 The radiation chart for the number of live births

Figure 3.4는 연도별 산모연령에 따른 출생아수의 분포를 선그래프와 삼차원 선그래프로 표시한 것이

다. Figure 3.4의 위쪽 선그래프를 보면 가장 안쪽 선그래프인 1981년의 출생아수의 분포모양은 산모연령 25세를 중심으로 가장 높이 분포하면서 오른쪽으로 길게 뻗은 분포모양을 하고 있고, 그 봉우리들이 점점 낮아져서 2106년에 이르러서는 가장 낮은 봉우리를 나타내고 있고, 중심위치가 산모연령 32세로 옮겨 감을 알 수 있다. 1981년부터 2016년 36년 동안 중심위치가 25세에서 36세로 11세 간격 이동한 것을 고려할 때 1년에 0.31세씩 산모연령의 중심위치가 변해왔음을 알 수 있다. 그러므로 향후 4 5년 후에 산모연령 중심위치는 33세로 바뀔 것이고 그 높이는 더 낮아질 것이며, 분포의 모양 역시 보다 더 좌우대칭형으로 변할 것이다 예측할 수 있다. 이러한 관점에서 탐색적 시각화 분석은 보다 쉽게 직관적인 예측결과를 제시한다.

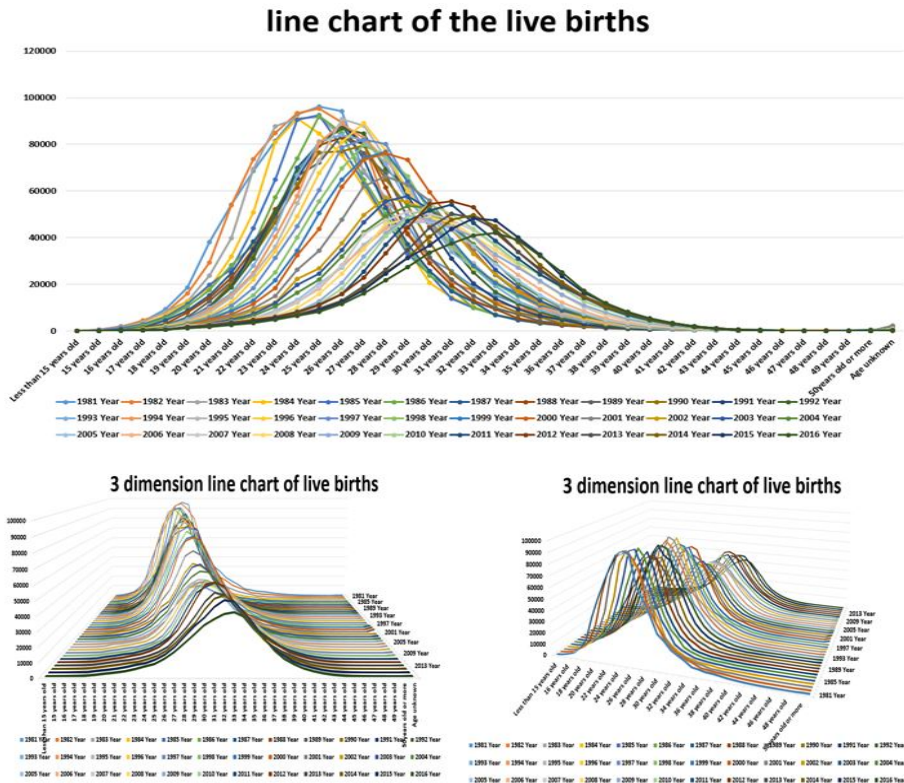


Figure 3.4 The line charts for the number of live births

Figure 3.5는 산모연령과 년도에 따른 출생아수의 분포를 하나의 그래프로 표현하였다. 그림의 왼쪽에 표시된 데이터 레벨들은 1981년도 산모연령에 따른 출생아수를 표시하였고, 그림의 오른쪽에 표시된 데이터 레벨들은 2016년 산모연령에 따른 출생아수를 표시하였다. 1981년에는 산모 나이의 25세에서 출생아수가 가장 최대를 나타내고, 2016년에는 32세에서 출생아수가 가장 최대를 나타낸다. 15세 미만에서 50세 이상까지의 좌표축 기준으로 볼 때, 1981년 오른쪽 긴 꼬리 모양에서 2016년으로 옮겨오에 따라 좌우대칭형으로 변해가는 모습으로 표현된다. 오히려 왼쪽 긴 꼬리 모양으로 변해 가는 양상을 보인다. 미래에는 33세 이상에서 출생아수의 최대를 나타내며, 출생아수는 조금씩 줄어들 것으로 예측된다.

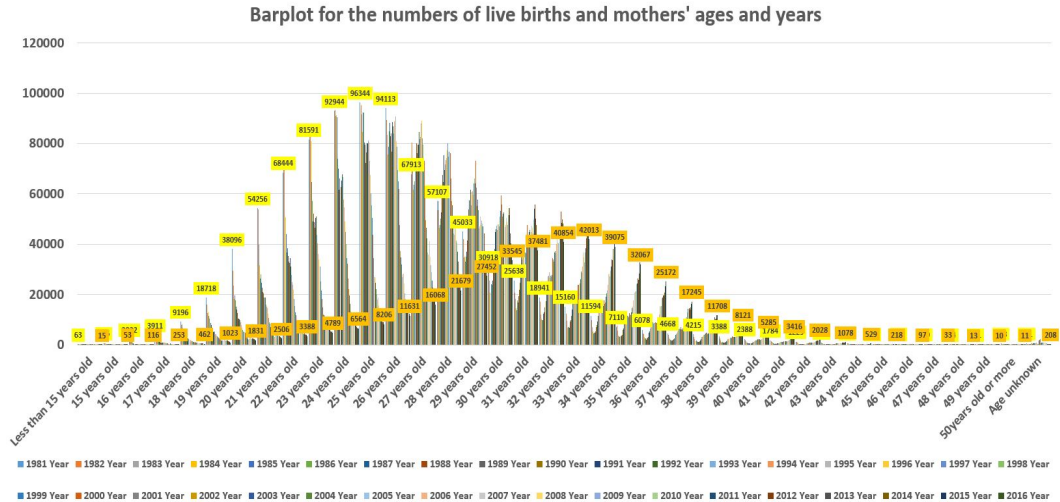


Figure 3.6는 연도에 따른 산모연령별 출생아수의 분포를 하나의 그래프로 표현하였다. 그림의 왼쪽 위쪽에 표시된 데이터 레벨들은 산모연령 25세에 따른 출생아수를 표시한 것으로 1981년에 산모연령 25세에 따른 출생아수가 최대를 나타내었지만 1994년을 기준으로 2016년까지 25세 산모들의 출생아수는 점점 감소하는 것을 보이고 있다. 반면에 산모연령 33세의 출생아수는 1985년부터 서서히 증가하기 시작하여 2003년에는 25세 산모들의 출생아수를 추월하여 점점 증가하는 현상을 보인다.

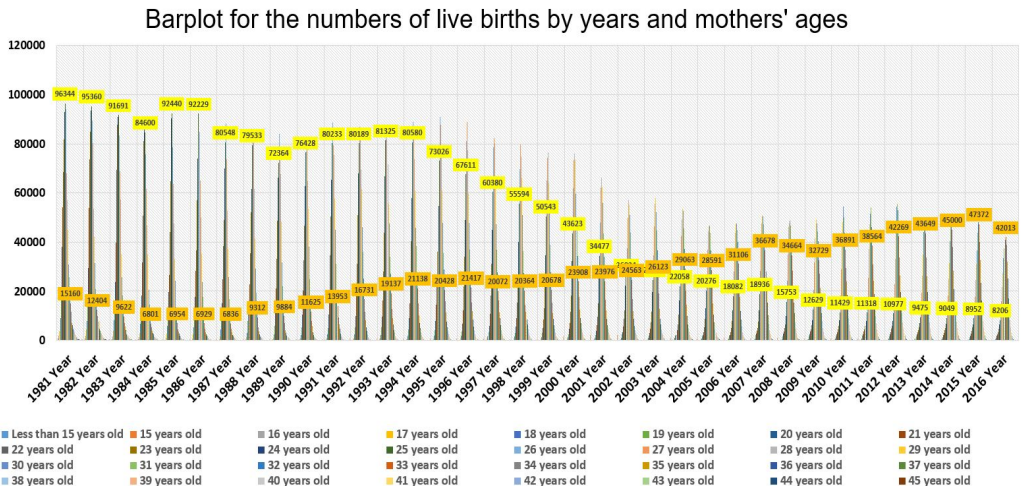


Figure 3.7은 년도에 따른 산모연령별 출생아수의 분포를 각각의 막대그래프로 표현하였다. 그림의 왼쪽 위쪽의 막대그래프로부터 연도가 바뀔에 따라서 점점 높이가 낮아지고 오른쪽 긴 꼬리모양에서 좌우대칭형으로, 좌우대칭형에서 왼쪽 긴 꼬리모양으로 변화하는 것을 볼 수 있다.

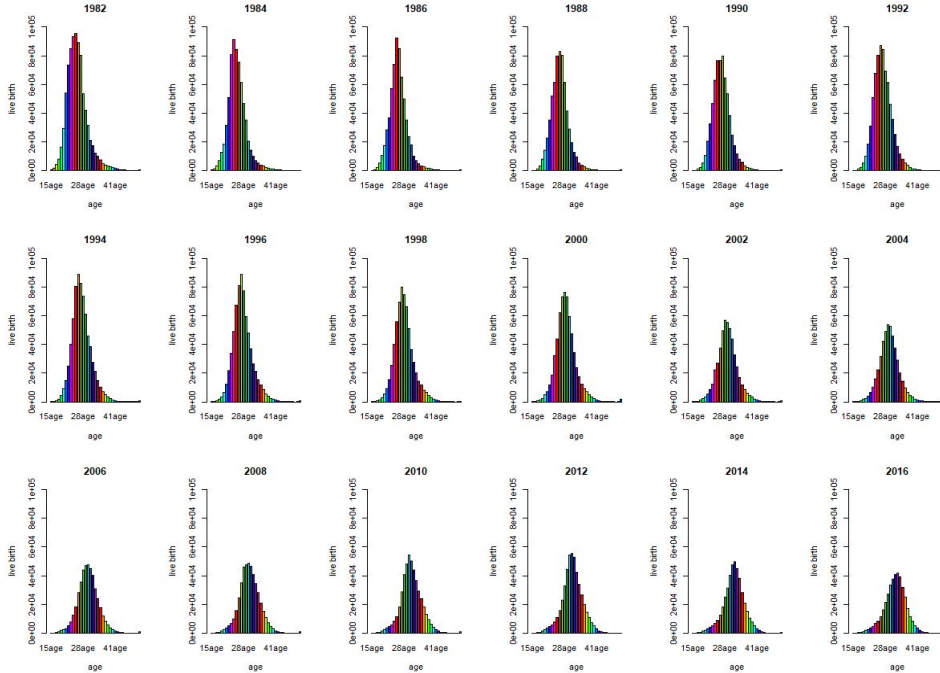


Figure 3.7 The barplots for the number of live births by years and mothers' age

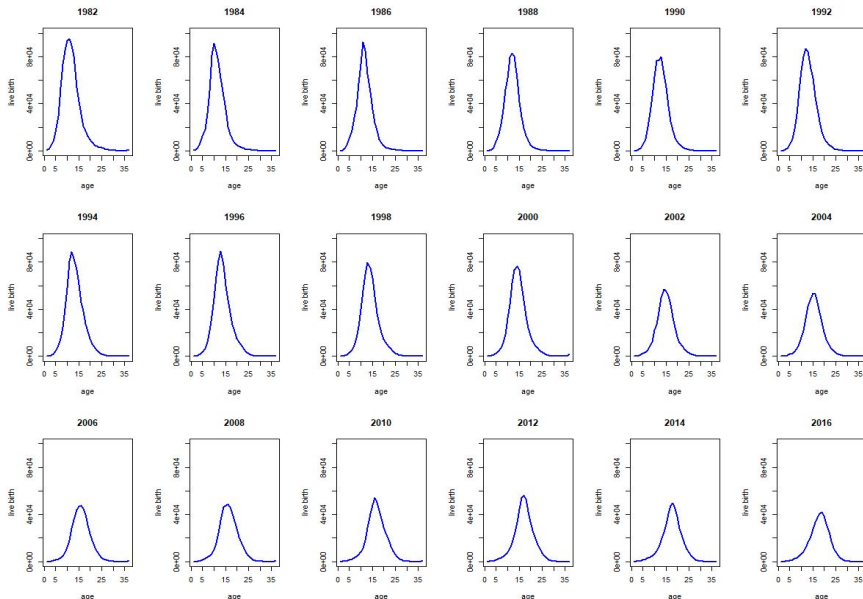


Figure 3.8 The time series plots for the number of live births by years and mothers' age

Figure 3.8은 년도에 따른 산모연령별 출생아수의 분포를 각각의 시계열그래프로 표현하였다. 그림의 왼쪽 위쪽의 막대그래프로부터 연도가 바뀌에 따라서 그래프 봉우리의 높이가 점점 낮아지고 오른쪽 긴 꼬리모양에서 좌우대칭형으로, 좌우대칭형에서 왼쪽 긴 꼬리모양으로 변해가는 것을 볼 수 있다.

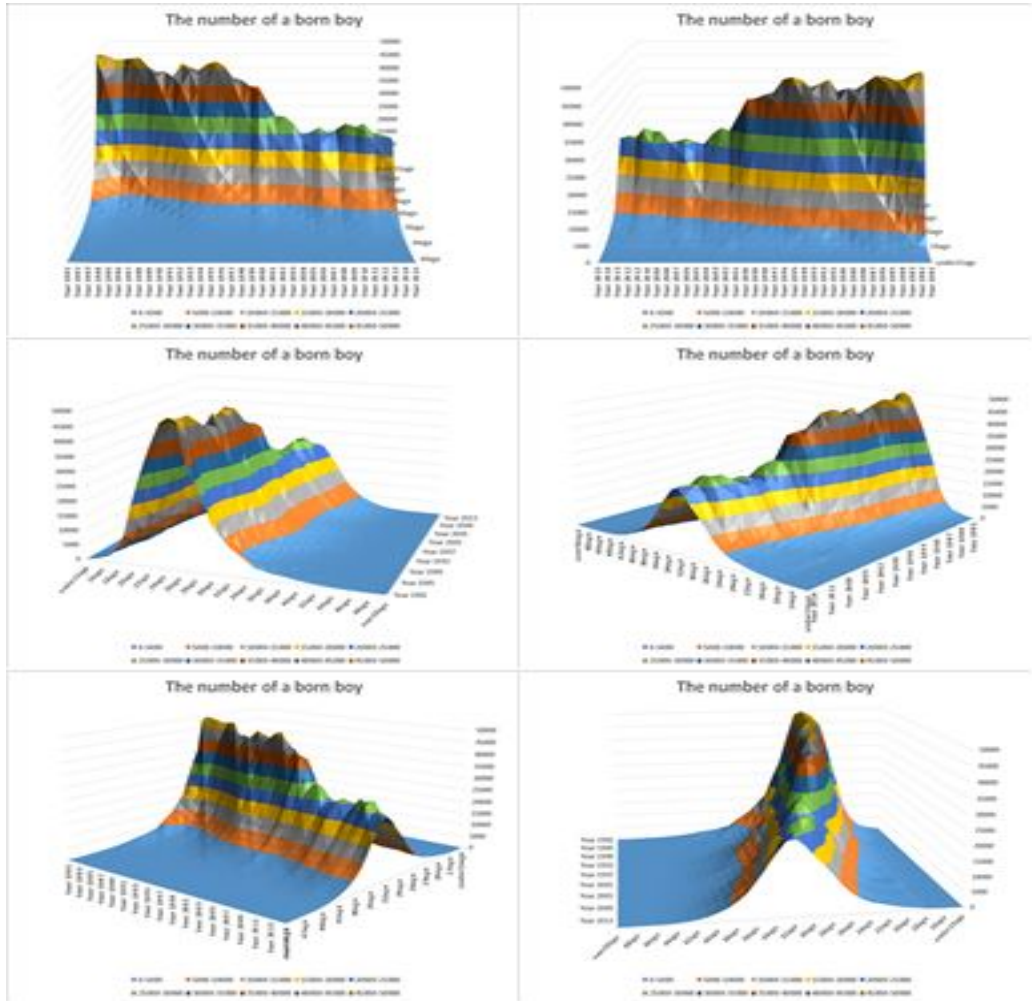


Figure 3.9 3-dimensional chart for the number of live births by boys

Figure 3.9은 년도에 따른 산모연령별 남자아이 출생아수의 분포를 삼차원 그래프들로 표현하였다. x -축의 좌표를 바꿈에 따라서, 삼차원 그래프의 각도가 바뀔에 따라서 근거리에는 분포의 모양과 특징들이 잘 설명되어 진다.

Figure 3.10은 기타의 삼차원 그래프들을 표현하였다. 여러 가지 다양한 시각화 그래들을 이용하여 출생아수 데이터를 다양하게 그들 분포의 모양과 특징들이 설명하고 있다.

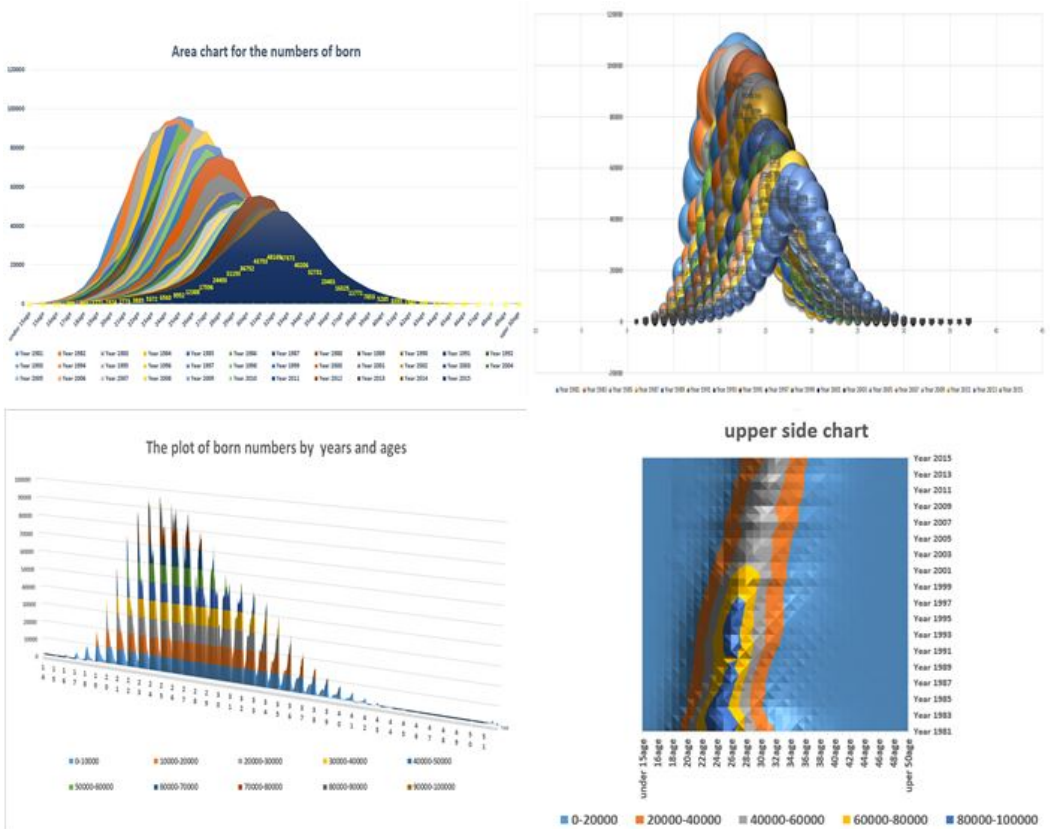


Figure 3.10 3-dimensional chart for the number of live births by boys

4. 각세별 산모연령별 출생아수의 분석과 결론

출생아수에 대한 하나의 데이터를 가지고 여러 가지 관점에서 시각화를 시도하였다, 하나의 데이터를 가지고 매우 많은 그래프 도구들로부터 다양한 분석을 도출할 수 있다는 것은 경이로운 것이다. 이는 데이터를 표현하는 관점에 따라서 매우 다양한 해석들을 내릴 수 있음이 보여 진다. 시각화 분석은 데이터가 가지고 있는 혹은 데이터가 스스로를 표현하고자 하는 속성들을 탐색적인 방법에서 혹은 교육적인 관점에서 쉽게 접근하고, 분석하고 예측할 수 있는 유용한 분석임에는 틀림이 없다.

본 논문에서는 정지된 시각화에서만 논의를 하였지만 동적 시각화 분석이 개발되고, 동적 시각화의 분석 방법이 다양해짐에 따라 데이터 분석에서의 새로운 시대가 열릴 수 있을 것이다. 더 나아가서 이러한 데이터 분석은 인공지능과 사물인터넷 등과 결합되어 한층 더 진화되는 분야가 될 것으로 생각된다.

Figure 4.1은 산모의 연령별 출생아수의 변화를 연도별로 나타낸 것이다. 산모의 연령에 따라서 출생아수의 패턴이 어떻게 달라지는지를 잘 나타내고 있다. 왼쪽 제일 위쪽의 산모연령 15세의 출생아수에 대한 그래프이고 2세씩 증가하여 오른쪽 가장 아래쪽 그래프가 산모연령 49세의 그래프이다.

이러한 데이터의 변화는 R-패키지 ‘changing point’를 사용하여 변화점 분석 (changing-point analy-

sis)을 시행하면 쉽게 그 위치 (변화점의 연도)를 추정할 수 있고, 그 추정된 변화점들을 기준으로 데이터를 나누어 선형모형론이나 시계열분석 등을 이용한 미래의 출생아수를 예측하는 것이 보다 오차가 적은 예측결과들을 제공해 줄 것이다. 변화점 추정과 예측에 대한 연구의 결과는 차후의 연구 논문에서 제시할 것이다.

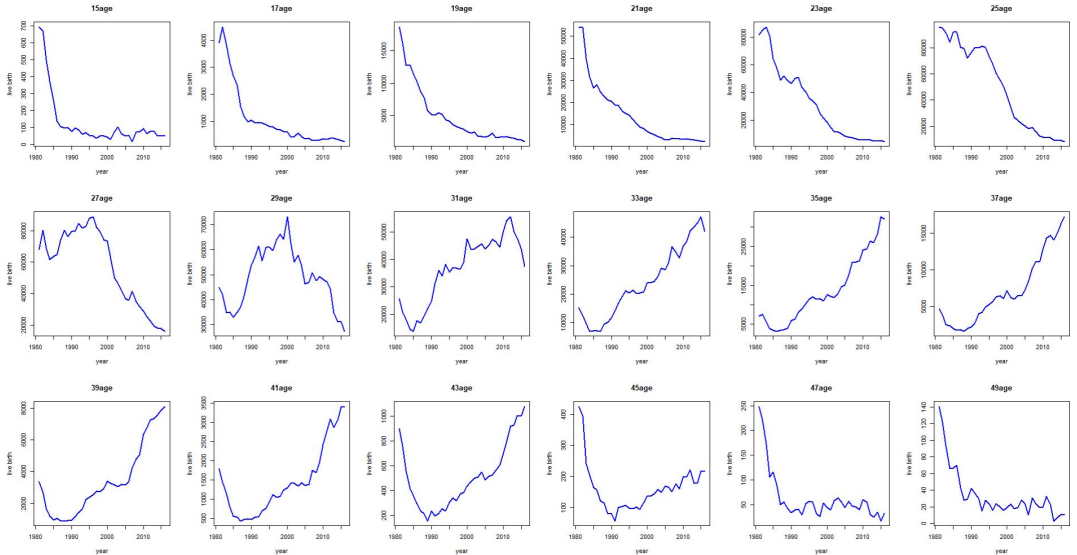


Figure 4.1 line charts for the number of live births by mother's ages

References

- Ahn, S., Lee, J. E. and Jang, D. (2017). Graphical method for evaluating the impact of influential observations in high-dimensional data. *Journal of the Korean Data & Information Science Society*, **28**, 1291-1300.
- Na, J. and Hwang, E. (2017). Graphical method for evaluating the impact of influential observations in high-dimensional data. *Journal of the Korean Data & Information Science Society*, **28**, 1279-1290.
- KOSIS (2017). *Live births by age group of mother, sex and birth Order period annual 1981 ~ 2016.*, http://kosis.kr/eng/statisticsList/statisticsList_01List.jsp?vwcd=MT_ETITLE\&parentId=A\#SubCont, Daejeon.
- Lee, G. (2017). Visualizing korean economic statistics. *Journal of the Korean Data & Information Science Society*, **28**, 1313-1326.
- Yoo, K. (2017). Effective visualization methods for a manufacturing big data system. *Journal of the Korean Data & Information Science Society*, **28**, 1301-1312.

Visualization analysis of the live births by mother's age[†]

Jongtae Kim¹

¹Department of Computing & Statistics, Daegu University

Received 20 December 2017, revised 9 January 2018, accepted 9 January 2018

Abstract

The purpose of this study is to analyze the statistical distribution of the number of births by visualization analysis using various visualization tools with data. And this data is the ‘live births by age group of mother, sex and birth order period annual 1981 ~ 2016’ on the Korean Statistical Information Service. In this study, we implemented the exploratory statistical analysis methods using R software and Microsoft Excel. We analyzed the changes in the number of births according to the changes of each age by visualization tools using various kinds of two-dimensional graphs such as pie graph, bar graph and line graph. And then the distribution of births was examined with visualization analysis using various three-dimensional graphs. Through the analysis of various exploratory visualizations, it is possible to predict the distribution of birth data and the number of births according to birth order or mother's age.

Keywords: Korean statistical information service, live births, visualization analysis.

[†] This papper was supported by research fund, Daegu University, 2015.

¹ Professor, Department of Statistics and Computer Science, Daegu University, Gyeongbuk 38453, Korea.
E-mail: jtkim@daegu.ac.kr